Razonamiento jurídico y ciencias cognitivas

Federico José Arena / Pau Luque / Diego Moreno Cruz
EDITORES

SERIE INTERMEDIA DE TEORÍA JURÍDICA Y FILOSOFÍA DEL DERECHO N30

Centro de Investigación en Filosofía y Derecho

Universidad Externado de Colombia

135 Años

RAZONAMIENTO JURÍDICO Y CIENCIAS COGNITIVAS

SERIE INTERMEDIA DE TEORÍA JURÍDICA Y FILOSOFÍA DEL DERECHO N.º 30

FEDERICO JOSÉ ARENA
PAU LUQUE
DIEGO MORENO CRUZ
(Editores)

RAZONAMIENTO JURÍDICO Y CIENCIAS COGNITIVAS

UNIVERSIDAD EXTERNADO DE COLOMBIA

ISBN 978-958-790-641-7

© 2021, FEDERICO JOSÉ ARENA, PAU LUQUE, DIEGO MORENO CRUZ (EDS.)
© 2021, UNIVERSIDAD EXTERNADO DE COLOMBIA

Calle 12 n.º 1-17 Este, Bogotá
Teléfono (57 1) 342 0288
publicaciones@uexternado.edu.co
www.uexternado.edu.co

Primera edición: julio de 2021

Diseño de cubierta: Departamento de Publicaciones
Corrección de estilo: Santiago Perea Latorre
Composición: Precolombi EU-David Reyes
Impresión y encuadernación:
Tiraje de 1 a 1.000 ejemplares

Impreso en Colombia
Printed in Colombia

FEDERICO JOSÉ ARENA
SEBASTIÁN FIGUEROA RUBIO
DANIEL GONZÁLEZ LAGIER
PAU LUQUE
MARÍA NATALIA ZAVADIVKER
MARÍA LAURA MANRIQUE
DANNY MARRERO AVENDAÑO
DIEGO MORENO CRUZ
ANDRÉS PÁEZ

CONTENIDO

Introducción.
Derecho, psicología y decisión judicial 11
Federico José Arena
Pau Luque
Diego Moreno Cruz

I. Filosofía y (neuro)ciencia: sobre la "naturalización"
de la filosofía práctica 23
Daniel González Lagier

II. Acción, expresión y atribución
de estados mentales 53
Sebastián Figueroa Rubio

III. Acción intencional de interpretación judicial.
Casos normales y sorpresivos 101
Diego Moreno Cruz

IV. La Razón y la razón. Psicología moral
y argumentación judicial 129
Pau Luque

V. Acerca de la relevancia de las investigaciones
sobre sesgos implícitos para el control
de la decisión judicial 157
Federico José Arena

VI. Los sesgos cognitivos y la legitimidad racional
de las decisiones judiciales 187
Andrés Páez

VII. Un argumento naturalista en contra
de la teoría retórica de la argumentación probatoria
en los procesos penales 223
Danny Marrero Avendaño

VIII. Entre el 'hombre promedio' y el 'hombre ideal': el criterio de razonabilidad y su aplicación al agente moral en el razonamiento cotidiano, jurídico y psicológico 247
María Natalia Zavadivker

IX. ¿Es el sometimiento de la víctima una alternativa al odio como estado mental? 279
María Laura Manrique

Introducción. Derecho, psicología y decisión judicial

FEDERICO JOSÉ ARENA
PAU LUQUE
DIEGO MORENO CRUZ

En las últimas décadas el estudio del razonamiento jurídico se ha visto enriquecido por las hipótesis que son formuladas a partir de los resultados de experimentos de la psicología cognitiva y de la neurociencia acerca de cómo el aparato cognitivo humano funciona y de qué modo las actividades inconscientes y conscientes de nuestra mente y cerebro influencian en nuestras acciones cuando lidiamos con cuestiones normativas. La filosofía experimental del derecho (*Experimental Jurisprudence*) es un modo de abordar problemas de la teoría general y la filosofía del derecho, que parte de la base de experimentaciones a partir de las cuales se sostiene que nuestras acciones están influenciadas por un conjunto de estados y operaciones mentales inconscientes y también por ideas normativas acerca de cómo debemos actuar en situaciones dadas. La filosofía experimental, entendida de este modo, reta a los teóricos y los filósofos del derecho a que formulen hipótesis que estén informadas por las explicaciones y predicciones de las ciencias cognitivas. Hipótesis que dejan en claro que el comportamiento de las personas está moldeado tanto por ideales normativos que ellas apoyan conscientemente como por tendencias psicológicas o procesos inconscientes que se llevan a cabo en su mente sin intención deliberativa alguna (SHAFIR, 2004: ix).

Por un lado, tales investigaciones han permitido poner bajo una nueva luz cuestiones tradicionalmente concebidas como puramente conceptuales o metafísicas. En este sentido, parte de la génesis de la idea de este libro se vincula con un intento por mostrar diferentes modos en los que sería posible enfrentar un reciente desafío a la filosofía analítica del derecho. La corriente tradicional de la filosofía analítica concebía al método filosófico como la reconstrucción racional de conceptos basada en las intuiciones del propio filósofo y en las propiedades generalmente asociadas con el concepto[1]. Actualmente, existe gran controversia acerca de cómo concebir la tarea del filósofo analítico y de la filosofía analítica del derecho en particular. La discusión se vincula con las objeciones avanzadas en contra de la distinción analítico-sintético. Según el punto de vista tradicional, los enunciados analíticos son aquellos cuyo valor de verdad depende

1 Este es el modo en que viene presentado, por ejemplo, por STRAWSON, 1992, y VON WRIGHT, 1963.

exclusivamente de su forma lógica y/o del significado de los términos que los componen. Los enunciados sintéticos, en cambio, son aquellos cuyo valor de verdad depende de cómo es el mundo. Esta constituyó la piedra basal de las ciencias empíricas según el positivismo lógico de la primera parte del siglo XX. Mientras que la filosofía debía encargarse de los enunciados analíticos, las ciencias empíricas se ocuparían de los enunciados sintéticos. Sin embargo, esta distinción fue atacada desde distintos ángulos; quizás la crítica más devastadora fue la formulada por Quine (QUINE, 1953). Si bien resulta difícil resumir en pocas palabras esa crítica, puede bastar con formular la conclusión, que se extrae de su objeción, del modo siguiente: hasta ahora no han sido ofrecidas pruebas suficientes para sostener la existencia de una clase de enunciados inmune a los hechos. Si ello es así, entonces todos los enunciados enfrentan el tribunal de la experiencia. Este resultado parecería dejar sin tarea a la filosofía, puesto que ningún conocimiento confiable sobre los hechos (de los que ahora depende también la verdad de los enunciados antes considerados analíticos) puede ser obtenido únicamente mediante el análisis conceptual; todos los enunciados han de ser testeados por la ciencia. Los filósofos analíticos, incluidos los filósofos del derecho, han enfrentado esta dificultad de distintas maneras[2].

En general puede afirmarse que existe un cierto acuerdo sobre que la filosofía analítica, al momento de teorizar, puede y debe ser sensible a la experiencia y receptiva hacia los resultados de las ciencias empíricas. Esto ha sido articulado de distintas maneras, la filosofía experimental es una de ellas. En sus inicios la filosofía experimental consistía en la aplicación de los métodos de la psicología experimental al estudio de las intuiciones[3]. En un sentido amplio, en cambio, la filosofía experimental comprende todo trabajo filosófico que usa métodos o resultados de investigaciones empíricas para enfrentar problemas filosóficos (KNOBE y NICHOLS, 2008).

Si bien no necesariamente todas las personas que participan en este libro comparten las tesis de la filosofía experimental, creemos que sus aportes presentan diferentes modos de extender este enfoque al ámbito

---

2 Entre muchos otros, pueden verse, DICKSON, 2001, y LEITER, 2007.

3 Algunos ejemplos son MCCANN, 2005: 737-748, y WEINBERG, NICHOLS y STICH, 2001: 429-460.

jurídico. Particularmente, sobre la base de avances acerca de cómo funciona realmente nuestro aparato cognitivo y cómo este influencia en la acción humana, se abordarán preguntas como: ¿en qué consiste atribuir un estado mental?, ¿es posible reducir las normas a estados de nuestra mente o a sucesos físicos del cerebro? Por otro lado, las investigaciones en ese ámbito han abierto nuevas líneas de investigación para el análisis del modo en que decidimos sobre cuestiones normativas. En particular, se presentan ahora preguntas como: ¿de qué modo las decisiones jurídicas son dependientes de la mente de los decisores?, ¿qué incidencia tienen los estados mentales y las emociones en el modo en que tomamos decisiones normativas?, ¿qué relación existe entre las actividades cognitivas de nuestra mente, los sucesos físicos y químicos de nuestro cerebro y la toma de decisiones racionales?

Los escritos que aquí se presentan y que componen el presente volumen tienen por objeto común el análisis de problemas de la filosofía del derecho y del razonamiento práctico de las personas (en especial, jueces y funcionarios) en relación con las normas jurídicas.

El conjunto de avances científicos apenas citados permite enfocar desde nuevos ángulos cuestiones y disputas clásicas de la teoría del derecho: ¿es el derecho en última instancia el conjunto de estados mentales o cerebrales de los que interpretan y aplican las disposiciones jurídicas? ¿O, en cambio, hay algo independiente de los estados mentales de los jueces, una suerte de reino de lo normativo, que establece qué es lo correcto, jurídicamente hablando, y que permite decir que los jueces, por lo menos en algunas ocasiones, pueden cometer errores? ¿Qué incidencia tienen las emociones de los jueces al momento de decidir casos concretos? ¿Es posible escindir la búsqueda racional de una respuesta jurídica correcta de las emociones que el caso provoca en el juez? ¿Los juicios normativos acerca de la corrección de las acciones propias y de otros pueden ser reducidos a la descripción de los sucesos físicos y químicos del estado en que se encuentra nuestro cerebro?

Los escritos abordan problemas relacionados con estas cuestiones, es decir, problemas que no son solo de teoría general del derecho, sino también de teoría de la interpretación, teoría de la responsabilidad, teoría del proceso y teoría de la prueba, entre otros.

Los dos primeros capítulos están dedicados a qué es lo que puede hacer la neurociencia en relación con los problemas de la normatividad del derecho. El tercer capítulo explora la interpretación judicial desde una

perspectiva psicologista de análisis. El cuarto capítulo aborda un debate metateórico desde una perspectiva de la psicología moral experimental. Los capítulos quinto, sexto y séptimo se ocupan de los sesgos implícitos y de su influencia en la toma de decisiones racionales de los jueces. El primero de estos capítulos lo hace en relación con el razonamiento interpretativo y los dos restantes en relación con el razonamiento probatorio. El octavo capítulo aborda el uso de criterios morales implícitos (razonabilidad y normalidad) en el juicio de corrección de las acciones de otros. Por último, el noveno examina la justificación de la agravación de la pena en relación con acciones realizadas por odio, entendido como motivo emocional.

En el primer capítulo, "Filosofía y (neuro)ciencia: sobre la 'naturalización' de la filosofía práctica", Daniel González Lagier plantea una serie de problemas en torno a la posibilidad de "naturalizar" la filosofía práctica, en concreto, la filosofía de la responsabilidad, sobre la base de explicaciones provenientes de la neurociencia: es decir, el intento de reemplazar las explicaciones de la filosofía sobre el razonamiento práctico-normativo y la acción humana con las explicaciones neurocientíficas acerca del funcionamiento del cerebro y de la formación de estados mentales. En especial, el autor problematiza la pretensión de la neuroética y del neuroderecho de naturalizar las normas y nuestros juicios normativos (morales) y la de naturalizar la mente (por ejemplo, reducir los estados mentales a estados cerebrales). Al respecto, concluye que la pretensión de reemplazar las explicaciones de la filosofía práctica con las explicaciones neurocientíficas es irrazonable. Sin embargo, también concluye que la naturalización es necesaria si con ella se entiende la pretensión de complementar, y no de reemplazar, las primeras explicaciones con las segundas: si bien las explicaciones ofrecidas por la filosofía son irreducibles a las explicaciones de las ciencias empíricas, aquellas no pueden contradecir la evidencia acerca de la naturaleza humana aportada por las segundas, por ejemplo, las explicaciones ofrecidas por la neurociencia.

En el segundo capítulo, "Acción, expresión y atribución de estados mentales", Sebastián Figueroa Rubio ofrece una revisión crítica de los supuestos de la neurociencia cognitiva en los que se basan las tesis, populares hoy en el estudio del derecho, a saber, el neuroderecho: la explicación de la mente, de la agencia y del comportamiento humano pueden, o tendrían que, traducirse en términos de estados y procesos cerebrales:

lo mental es reducible a lo cerebral. El autor muestra cómo la aceptación de tales supuestos repercute no solo en la comprensión de la mente y de su relación con el comportamiento humano, sino en nuestra comprensión de la práctica de atribución de estados mentales en el ámbito jurídico. Por último, Figueroa defiende una noción diferente de agencia, y de la mente, que él presenta como una representación no-reduccionista, que además es más adecuada para dar cuenta de la práctica de atribuir estados mentales a otros en el contexto del derecho: nuestro comportamiento es la expresión de lo mental y las acciones son adscritas a otros en relación con un contexto particular. Si bien un estado mental presupone un estado cerebral, es decir, hay una correlación entre ambos tipos de estados, lo conceptual no es reducible a lo empírico. El neuroderecho, por tanto, se equivoca al asimilar, por ejemplo, el dolo con un estado físico en el que se encuentra el cerebro.

En el tercer capítulo, "Acción intencional de interpretación judicial. Casos normales y sorpresivos", Diego Moreno Cruz defiende un abordaje psicologista de la interpretación judicial, presentándola como una actividad de formación de la intención del juez de proferir actos de habla interpretativos. Y señala que ella algunas veces puede consistir en un mecanismo intuitivo, frente a los casos normales, y otras veces en un proceso de decisión, frente a los casos sorpresivos. El autor cuestiona que haya casos respecto de los que los jueces no ejercen discrecionalidad para obtener la norma jurídica que les es aplicable. Para ello, ofrece una distinción entre dos tipos de análisis psicologista: prospectivo y retrospectivo. Y afirma que, aunque la identificación de normas jurídicas aplicables a los casos normales no requiera de la toma de decisión interpretativa en un sentido prospectivo de análisis de la interpretación judicial, toda identificación de normas jurídicas aplicables a los casos, ya sea normales o sorpresivos, involucra el ejercicio de la discrecionalidad en un sentido retrospectivo.

En el cuarto capítulo, "La Razón y la razón. Psicología moral y argumentación judicial", Pau Luque expone una posible conexión entre la teoría de la argumentación jurídica, en concreto, judicial, y la psicología moral experimental. De modo específico, con base en la literatura en torno a esta última el autor plantea una crítica a una noción de argumentación jurídica sostenida por la escuela de la Universidad de Alicante. Además, el autor recomienda una noción de razonamiento que tome en cuenta la evidencia

empírica ofrecida por las ciencias cognitivas, para obtener lo que según él sería una buena teoría de la argumentación jurídica. En este sentido, él desacuerda con la noción de argumentación que considera que el juez encuentra una razón para decidir un caso correctamente, de tal manera que tal razón no solo explica su decisión sino que también la justifica. Por su parte, da razones para considerar que la razón que justifica la decisión correcta no explica el porqué el juez tomó una decisión: la razón no es causa de la decisión sino tan solo su justificación.

En el quinto capítulo, "Acerca de la relevancia de las investigaciones sobre sesgos implícitos para el control de la decisión judicial", Federico José Arena se ocupa de exponer las investigaciones en torno a los sesgos cognitivos implícitos para el control de la decisión judicial. El autor afirma que los sesgos implícitos entendidos como causa de las acciones que realizan los jueces para solucionar los casos requieren de un criterio normativo para ser identificados. Sin embargo, muestra que si tales criterios jurídicos existen, los sesgos son irrelevantes. Si tales criterios no existen, entonces se acudirá a criterios extrajurídicos que a su vez serán suficientes para determinar la corrección de la decisión del juez, y los sesgos, también en estos casos, serán irrelevantes para este fin. No obstante, el autor sostiene que los sesgos implícitos, si bien pueden ser irrelevantes en las decisiones individuales tomadas por cada juez, pueden ser identificados al comparar dos o más decisiones individuales sobre casos similares, lo cual exigiría adoptar una perspectiva estructural para su abordaje.

En el sexto capítulo "Los sesgos cognitivos y la legitimidad racional de las decisiones judiciales", Andrés Páez expone una serie de sesgos implícitos producidos por diferentes heurísticas en la actividad decisional de los jueces y ofrece una propuesta para minimizar los efectos perjudiciales que se atan a ellos, principalmente el efecto negativo de que la decisión del juez pueda considerarse irracional y por lo tanto ilegítima. Respecto de la autoevaluación del sistema judicial y de las decisiones de los jueces, el autor considera que la minimización de los efectos de los sesgos cognitivos y un mayor grado de racionalidad en sus decisiones pueden obtenerse por medio de la formulación de guías en la toma de decisión de los jueces que resulte en una protocolización del actuar procesal. La idea general sobre la que opera el argumento de Páez es que, en lugar de modificar los comportamientos de los agentes judiciales, se podría alterar el entorno

dentro del cual aquellos toman decisiones para reducir la probabilidad de que estos sesgos aparezcan.

En el capítulo séptimo, "Un argumento naturalista en contra de la teoría retórica de la argumentación probatoria en los procesos penales", Danny Marrero Avendaño examina el efecto que el sesgo confirmatorio tiene en la configuración de los hechos relevantes en los procesos penales. En especial en las creencias y actitudes de los fiscales que buscan reforzar la evidencia de inculpación y minimizar la evidencia exculpatoria en relación con el procesado. El autor expone una teoría retórica de la argumentación probatoria (entendida como un proceso de inferencias) y ofrece razones para considerarla inadecuada para llevar a cabo la finalidad de dar un trato sustancialmente justo a las personas involucradas. Y propone, para su reemplazo, una teoría epistemológica de la argumentación que esté dirigida a reducir los sesgos confirmatorios en aras de que la investigación penal, que tiene por propósito reconstruir los hechos objeto de la disputa penal, sea lo más imparcial posible.

En el capítulo octavo, "Entre el 'hombre promedio' y el 'hombre ideal': el criterio de razonabilidad y su aplicación al agente moral en el razonamiento cotidiano, jurídico y psicológico", María Natalia Zavadivker aborda el tema de los criterios de corrección moral usados para juzgar las acciones de otros, considerando cuestiones acerca de cómo se comportan y cómo se deberían comportar los agentes morales. Zavadivker toma como objeto de su análisis la aplicación del criterio de razonabilidad y los estándares sobre los que este se apoya en el ámbito jurídico, por ejemplo, el del "hombre promedio". Analiza distintas nociones de "normalidad" y las relaciones que estas tienen con el criterio de razonabilidad. Muestra cómo muchas veces se confunden los planos diferentes en los que son aplicables la normalidad y la razonabilidad. En este sentido, sostiene que los juicios de corrección moral presuponen estándares de normalidad y, por lo tanto, la empresa descriptiva de los psicólogos morales es dependiente de una evaluación moral implícita que está basada en un criterio normativo a partir del cual es posible identificar desviaciones en la acción de las personas.

En el noveno y último capítulo, "¿Es el sometimiento de la víctima una alternativa al odio como estado mental?", María Laura Manrique se ocupa del modo en que las emociones motivan la acción de las personas y del rol que estas juegan en la atribución de responsabilidad penal. De

modo específico, la autora aborda el tema del odio como una emoción motivacional del agente, y también como una justificación de agravación de la pena. Al respecto, la autora expone una posición dogmática según la cual la justificación de la agravación por cometer ilícitos por odio está en que el victimario tiene por pretensión someter a la víctima. En contraposición, ella considera que es más fructífero abordar la justificación de la agravación, con fines explicativos y justificativos, considerando al odio como un estado mental del victimario. Y afirma que este modo de concebir al odio, como estado mental, es además necesario para sostener la posición a la que ella dirige sus objeciones, es decir, al odio como sometimiento.

Es importante que, en general, la filosofía del derecho saque provecho de las investigaciones en psicología experimental y psicología cognitiva. Este es uno de los objetivos principales de este libro: cubrir un espacio de discusión en teoría del derecho que no ha sido prácticamente explorado en español. En este sentido, esta obra podría ser una suerte de libro fundacional del cruce de caminos entre la psicología y la filosofía del derecho en el mundo de habla hispana. El volumen está dirigido a todos aquellos filósofos del derecho, filósofos de la moral y, por supuesto, a todos aquellos interesados en el desarrollo de la psicología cognitiva y la psicología experimental. Pero también está dirigido a todos aquellos que ejercen el derecho no solo desde su vertiente teórica, sino también desde su vertiente práctica. El cruce de caminos entre la psicología, la teoría y la filosofía del derecho ayudará a entender cómo opera el derecho en los casos concretos. Por ello, abogados, jueces, fiscales y procuradores deberían encontrar enormemente útiles los debates filosóficos, teóricos y psicológicos que esta obra contiene.

REFERENCIAS

DICKSON, J., 2001: *Evaluation and Legal Theory*. Oxford-Portland: Hart.

KNOBE, J. y NICHOLS, S. (eds.), 2008: *Experimental Philosophy*. Oxford-New York: Oxford University Press.

LEITER, B., 2007: *Naturalizing Jurisprudence. Essays on American Legal Realism and Naturalism in Legal Philosophy*. Oxford: Oxford University Press.

McCann, H. J., 2005: "Intentional action and intending: recent empirical studies", *Philosophical Psychology*, 18: 737-748.

Quine, W. V. O., 1953: "Two dogmas of empiricism", en Quine, W. V. O., *From a Logical Point of View*. Cambridge: Harvard up.

Shafir, E., 2004: "Introduction and biography", en Tversky, A., *Preference, Belief, and Similarity,* Shafir, E. (ed.). Cambridge, Mass.: mit Press.

Strawson, P. F., 1992: *Analysis and Metaphysics. An Introduction to Philosophy*. New York: Oxford University Press.

Von Wright, G. H., 1963: *Norm and Action*. London: Routledge and Kegan Paul, 1963.

Weinberg, J. M., Nichols, S. y Stich, S., 2001: "Normativity and epistemic intuitions", *Philosophical Topics*, 29: 429-460.

I.
Filosofía y (neuro)ciencia: sobre la "naturalización" de la filosofía práctica

DANIEL GONZÁLEZ LAGIER*

La relación entre la filosofía y la ciencia es un viejo y perenne problema filosófico, que el reciente desarrollo de la neurociencia ha revitalizado, a veces con el rótulo de "naturalización de la filosofía". Puede afirmarse que la filosofía en torno a la responsabilidad, la filosofía práctica (que es la que aquí nos interesa), está experimentando un proceso de "naturalización", que es una consecuencia o manifestación de lo que se ha llamado "la naturalización de la filosofía" en general[1]. En este trabajo trataré de dar cuenta de algunos intentos y algunos problemas y límites de la "naturalización" de la filosofía práctica en términos de explicaciones neurocientíficas.

1. ¿Qué entienden los filósofos por "naturalizar" la filosofía?

No hay un claro acuerdo acerca de qué entender por "naturalización" de la filosofía, pero se podría coincidir en que implica la reconstrucción de la filosofía a partir de conceptos admitidos –o, al menos, admisibles– por las ciencias de la naturaleza o ciencias empíricas (frecuentemente identificadas, en un sentido estricto de "ciencia", con la física, la química o la biología[2]) (Moya, 2005: 59). Esta es una caracterización amplia (y mínima),

---

* Catedrático de Filosofía del Derecho de la Universidad de Alicante (España). Co-director de la colección "Derecho y Literatura" (Marcial Pons). Co-secretario de edición de "Doxa. Cuadernos de Filosofía del Derecho". Profesor en el máster de Argumentación Jurídica de Alicante y en el máster de Razonamiento Probatorio de la Universidad de Girona (España). Investigador de la Cátedra de Cultura Jurídica de la Universidad de Girona (España).

1 Para un análisis de este fenómeno puede verse Grimaldos y Pacho, 2005.

2 El impulso que se encuentra muchas veces detrás de estas concepciones es el fisicalismo, esto es, la tesis de que todo lo que existe es, en última instancia, físico y de que todo fenómeno puede ser explicado, en última instancia, en términos de propiedades físicas. A veces se habla de naturalización de la ética, de la filosofía del derecho o de la teoría del derecho no en términos de reducción a una ciencia "básica" como la física, sino a ciencias como la sociología o a una psicología no fundada neurocientíficamente. No me referiré aquí a este sentido de "naturalización", más débil y menos novedoso (desde esta perspectiva, el utilitarismo es una moral naturalizada y el realismo jurídico una teoría del derecho naturalizada, pero esta reducción no sería suficiente para un reduccionista "fuerte").

que abarca tanto las posturas más cientifistas, que pretenden la reducción completa de la filosofía (o una parte de la misma) a alguna ciencia empírica (como propuso Quine a propósito de la epistemología, a la que consideraba psicología cognitiva [QUINE, 2002])[3], como las posturas menos radicales, que entienden que hay cierto continuo entre filosofía y ciencia, aunque ambas puedan tener dominios distinguibles. A las primeras concepciones podemos llamarlas "tesis del reemplazo" (de la filosofía por la ciencia) y a las segundas "tesis de la complementariedad" (MARTÍNEZ y OLIVÉ, 1997: 16). El enorme desarrollo que las investigaciones sobre el funcionamiento del cerebro y el sistema nervioso está teniendo desde mediados del siglo XX, junto con el papel central de los procesos cerebrales en el razonamiento y la acción, explican que la neurociencia se haya convertido en el paradigma conceptual y metodológico al que se está intentando reconducir la filosofía práctica.

Suele situarse el nacimiento de la neurociencia moderna en 1888, cuando Santiago Ramón y Cajal descubre que el cerebro es una red de células, las neuronas. Posteriormente, Charles Sherrington analizó las conexiones entre ellas y Edgard Adrian registró actividad bioeléctrica en todo el sistema nervioso. Los avances en psicofarmacología a mediados del siglo XX y, sobre todo, la aparición de las técnicas de neuroimagen en los noventa y principios del siglo XXI han dado un impulso decisivo a la neurociencia. La gran mayoría de noticias espectaculares que trascienden a la opinión pública tienen que ver con estas técnicas de neuroimagen, que han dado lugar a lo que se ha llamado la neurociencia cognitiva, esto es, el estudio del funcionamiento del cerebro en los procesos de adquisición del conocimiento y de la formación de estados mentales[4]. Del tronco central de la neurociencia se han ramificado nuevas disciplinas que pueden verse como aplicaciones de la misma a distintos ámbitos. En 2002 se celebró en San Francisco el congreso "Neuroética: esbozando un mapa del terreno", que

---

3 Este artículo puede verse como uno de los más influyentes a favor de la naturalización de la filosofía.

4 Una descripción detallada de la evolución del conocimiento del cerebro desde el Antiguo Egipto puede verse en CHAUNGEUX, 1985. Véase también GIMÉNEZ ANAYA y MURILLO, 2007.

consagró esta nueva disciplina, dedicada a la aplicación de la neurociencia a los temas tradicionales de la ética, y en 2007 la MacArthur Foundation creó el proyecto "Derecho y Neurociencia" para reunir a varios conocidos neurocientíficos, filósofos y juristas de diversos países con el fin de profundizar en la intersección entre las neurociencias y el derecho.

En realidad, hay dos maneras de entender la "neuroética" y el "neuroderecho" (ROSKIES, 2002; BONETE PERALES, 2010: 64; CORTINA, 2010: 131): (1) como *ética (o derecho) de la neurociencia*, esto es, una parte de la ética (o del derecho) que trataría de establecer un marco ético (o jurídico) para las investigaciones neurocientíficas y sus aplicaciones; y (2) como *neurociencia de la ética (o del derecho)*, esto es, el estudio de la conducta ética (o de problemas jurídicos) desde el punto de vista de las investigaciones sobre el cerebro. Algunos problemas de la ética y el derecho de la neurociencia son los siguientes: si está justificado o no el uso de los descubrimientos neurocientíficos para la mejora de las capacidades mentales o sensoriales de los humanos (el llamado "transhumanismo"), en qué condiciones es legítimo el uso en los tribunales de pruebas basadas en técnicas neurocientíficas (como la prueba P300 o *brainfingerprinting*, que se espera que permita determinar si el sujeto miente observando las variaciones en las ondas cerebrales ante ciertos estímulos), qué valor en relación con la atribución de responsabilidad hay que conceder a determinadas disfunciones cerebrales, o si es correcto –y en qué casos– usar técnicas de control de la conducta basadas en conocimientos neurocientíficos. Y algunos problemas de la neurociencia de la ética y del derecho son los siguientes: la discusión general sobre el libre albedrío y su relación con la responsabilidad, la cuestión de la fundamentación de las normas penales en emociones como la repugnancia, el análisis del papel de la oxitocina o de las llamadas "neuronas espejo" en nuestra conducta ética, o si la neurociencia puede fundamentar conclusiones normativas acerca de la corrección de nuestros juicios morales (o nuestras decisiones jurídicas).

La "neuroética" y el "neuroderecho", entendidos como neurociencia de la ética y neurociencia del derecho, pueden verse como el principal intento actual de naturalizar la filosofía práctica y la responsabilidad (si bien el ámbito de la filosofía práctica puede considerarse más amplio que el de la "filosofía de la responsabilidad", asumiré que esta es su núcleo central).

2. LAS DOS VÍAS DE "NATURALIZACIÓN" DE LA FILOSOFÍA PRÁCTICA: LA NATURALIZACIÓN DE LAS NORMAS Y LA NATURALIZACIÓN DE LA MENTE

De acuerdo con la concepción clásica de la responsabilidad (estoy pensando en la responsabilidad moral y jurídico-penal, entendida como reproche), para adscribir responsabilidad a un agente es necesario –al menos– (1) que exista un sistema de reglas (con el que juzgamos la conducta del individuo) y (2) que el agente haya actuado libremente, en el doble sentido de que tenga *libertad de acción* (que nuestras acciones sean consecuencias de la combinación de nuestros deseos con las creencias acerca de cómo satisfacerlos) y *libertad de voluntad* (que esos deseos y creencias sean a su vez, al menos en cierto grado, libres, controlables por el agente)[5]. Consecuentemente, el proceso de fundamentación de la filosofía práctica a partir de las ciencias empíricas se está llevando a cabo por dos vías (siendo necesario recorrer las dos para el éxito completo del proyecto): la primera, la "naturalización de la normatividad"; la segunda, la "naturalización de la mente" (y, con ella, de la acción).

La naturalización de las normas es el intento de dar cuenta de la normatividad a partir de las ciencias empíricas, en la línea del "darwinismo moral" de Spencer, que trataba de fundamentar las normas éticas en el proceso evolutivo, o de la "sociobiología" de E. O. Wilson, que pretendía explicar el comportamiento de todos los animales extrapolando el egoísmo genético del mundo de los genes a los demás ámbitos de la vida. En la actualidad, desde la neuroética muchos autores (J. Haidt, M. Hauser, Patricia Churchland y M. S. Gazzaniga, entre otros) proponen dar cuenta de las opiniones y creencias morales como un conjunto de intuiciones, emociones y capacidades en gran parte innato, inscrito en nuestro cerebro por las "fuerzas" de la evolución:

> Hemos desarrollado –escribe Hauser– un instinto moral, una capacidad que surge en cada niño, diseñada para generar juicios inmediatos sobre lo

5 Asumo, de entrada, la idea tradicional de que la responsabilidad presupone la libertad, sin entrar en las discusiones entre compatibilistas e incompatibilistas.

> que está moralmente bien o mal, sobre la base de una gramática inconsciente de la acción. Una parte de esta maquinaria fue diseñada por la mano ciega de la selección darwiniana millones de años antes de que apareciera nuestra especie; otras partes se añadieron o perfeccionaron a lo largo de nuestra historia evolutiva y son exclusivas de los humanos y de nuestra psicología moral (HAUSER, 2008: 17)[6].

La naturalización de la mente, por su parte, es el intento de reconstruir los conceptos mentales –como creencia, decisión, intención, deseos, emociones, dolor, etc.– de manera que puedan ser aceptados por las ciencias de la naturaleza. La neurociencia ofrece técnicas –como, por ejemplo, la resonancia magnética funcional– que permiten detectar los cambios en el flujo sanguíneo en el cerebro en el momento en que el individuo realiza determinadas tareas motoras o está en ciertos estados cognitivos o emocionales, lo que parece permitir correlacionar estados mentales con estados cerebrales, de manera que ante la presencia de un estado cerebral determinado se podría suponer que el sujeto tiene el estado mental correspondiente. Esto sugiere una estrecha conexión entre estados mentales y estados cerebrales, conexión que los "neurofilósofos" interesados en la mente pueden tratar de usar para explicarla (veremos más adelante estos intentos de explicación de la mente).

En lo que sigue trataré de dar cuenta de algunos problemas de la naturalización de las normas y de la naturalización de la mente, que apuntan a la imposibilidad (e irrazonabilidad) de satisfacer la tesis del reemplazo, pero también a la necesidad de la tesis de la complementariedad.

3. ALGUNOS PROBLEMAS DE LA NATURALIZACIÓN DE LAS NORMAS

3.1. Autores como Kant, Hume (o cierta interpretación de este autor) y Moore han contribuido a que el mundo de las normas se haya concebido

---

6 Dado que una buena parte de los valores jurídicos suelen ser un reflejo de los valores morales de la sociedad, este proceso de naturalización afecta también al derecho, y no solo a la moral.

como un ámbito separado e independiente del de la naturaleza, pero la neurociencia ha aportado argumentos con los que se ha puesto en cuestión esta imagen de lo normativo, tratando de mostrar que la moral no tiene tal autonomía. La principal vía para esta "naturalización de la normatividad" tiene relación con ciertas conclusiones a las que se ha llegado a partir de la realización de "experimentos éticos" de tipo psicológico y neurocientífico. Jonathan Haidt, por ejemplo, ha realizado una gran cantidad de encuestas en las que planteaba ciertos problemas morales, llegando a la conclusión de que las respuestas se basaban en rápidas intuiciones que luego los sujetos no sabían cómo racionalizar. A partir de ahí propuso que la moral es más una cuestión de intuiciones y emociones que de razones (HAIDT, 2012). Posteriormente Joshua Greene examinó con técnicas de neuroimagen la actividad cerebral de sujetos a los que se les planteaban diversos dilemas morales, llegando a la conclusión de que en los problemas morales que afectaban más personalmente a los sujetos había una mayor actividad de la zona que regula las emociones (GREENE, 2012). Marc Hauser, también por medio de encuestas que apuntaban a rasgos morales universales, ha propuesto la hipótesis de la existencia de un órgano moral, esto es, de una capacidad innata para el desarrollo de códigos morales que determina en parte el contenido de estos (algo así como una gramática moral universal, en analogía con la gramática universal postulada por Chomsky) (HAUSER, 2008). En esta línea, William Casebeer ha afirmado que la teoría moral aristotélica de la virtud es más plausible desde un punto de vista neurobiológico que la teoría moral kantiana o la de John Stuart Mill.

El punto de partida de Casebeer es que cada una de estas teorías contiene implícitas una psicología moral específica que exige capacidades cognitivas diferentes. Así, la teoría de Kant parecería requerir "al menos la capacidad de comprobar la consistencia lógica de máximas universalizadas de una manera independiente de la contaminación del afecto y la emoción", capacidad que se correspondería con las funciones de la región frontal del cerebro. La teoría utilitarista de Mill requiere la capacidad de realizar cálculos utilitarios y el cultivo de emociones que nos muevan a procurar la felicidad de los demás, lo que implicaría las regiones prefrontal, límbica y sensorial del cerebro. La ética aristotélica de la virtud, por último, sería la más exigente, porque requiere educar nuestro carácter de manera que nuestros apetitos se coordinen con las buenas razones; esto

implica una "psicología global" que exige una intervención coordinada de las regiones del cerebro anteriormente mencionadas. Pues bien, nuestro autor cree que hay pruebas para aceptar, tentativamente, que la "cognición moral" pone en marcha de manera coordinada diferentes sistemas y redes cerebrales relacionadas tanto con la cognición como con las emociones (es decir, las regiones prefrontal, frontal, límbica y sensorial: lo que podría llamarse "la zona de la cognición moral"), lo cual muestra que "existe una clara convergencia entre la neuroética contemporánea y la psicología moral aristotélica" (Casebeer, 2003).

Como vemos, de lo que se trata es de derivar conclusiones acerca de la moral a partir de descripciones empíricas, de explicar la moral a partir del funcionamiento del cerebro. Sin embargo, algunas propuestas no se quedan en un nivel meramente descriptivo y explicativo, sino que tratan de derivar conclusiones normativas (como parece ser el caso de Casebeer). En sus versiones más radicales, se trata de reducir los valores o las normas morales a regularidades naturales, a los valores o las pautas de comportamiento que surgen de nuestra naturaleza física, asumiendo su carácter normativo y construyendo de esta forma un *realismo moral naturalizado.* De acuerdo con esto, la moral puede descubrirse por medios empíricos y las normas no forman parte de un reino distinto de la naturaleza.

Podemos caracterizar este "realismo moral neurocientífico" –que sostendrían autores como M. Hauser, Patricia Churchland (Churchland, 2012) o M. Gazzaniga (Gazzaniga, 2015), entre otros– a partir de los siguientes rasgos[7]:

A. Intuicionismo: nuestro juicio moral está dominado, o fuertemente influido, o condicionado en gran medida por intuiciones rápidas, inconscientes, no intencionales, ajenas a la reflexión. Ante los dilemas morales, nuestra respuesta suele ser intuitiva, irreflexiva, al menos en un primer momento. Por tanto, la deliberación y discusión racional sobre los problemas morales no tiene relevancia genuina (o la tiene muy menor, o solo en los casos en los que no nos vemos afectados personalmente). Su papel –en la mayoría de las ocasiones–, o bien es una racionalización *ex post*, o bien solo pretende causar en el oponente la misma intuición.

---

7 Para un análisis más detenido véase González Lagier, 2017.

B. Emotivismo: estas intuiciones morales dependen fundamentalmente de nuestras emociones o son la expresión de estas emociones, como demuestran las pruebas neurofisiológicas basadas en la observación de las zonas del cerebro que tienen mayor actividad en el momento en que nos enfrentamos a dilemas morales. La tesis de que las decisiones morales provienen de la emoción –manifestada como intuición– puede verse como un caso especial de la tesis general que enfatiza el papel de las emociones en la toma de decisiones (de cualquier tipo). Damasio sugiere que el modelo clásico de toma de decisiones como el resultado de una ponderación de razones a favor o en contra de una u otra respuesta no es descriptivamente adecuado. Los seres humanos no deciden en realidad de esa manera (si así fuera, las decisiones racionales, simplemente, no serían posibles, porque no es posible procesar toda la información necesaria en el tiempo del que disponemos para actuar). Las emociones nos ayudan a tomar decisiones racionales porque "filtran" las alternativas posibles en función de si en el pasado alternativas semejantes fueron "etiquetadas" como adecuadas (produjeron placer o satisfacción) o inadecuadas (produjeron frustración o dolor). Las emociones son, por tanto, "marcadores somáticos" de las posibles alternativas (DAMASIO, 2004: 201 y ss.).

C. Innatismo: se trata de intuiciones y emociones innatas, transmitidas genéticamente, aunque luego pueden ser moldeadas culturalmente: nacemos con valores, creencias o principios morales independientes en su origen del aprendizaje, de manera que la moral no es enteramente un producto cultural.

D. Evolucionismo: las intuiciones morales se ven como mecanismos que la evolución ha seleccionado porque aseguran la supervivencia de la especie. Uno de lo principios que estaría detrás de estas intuiciones, según muchos neuroéticos, sería el altruismo, la cooperación o el principio de beneficencia. Pues bien, en opinión de autores como Michel Ruse, el "altruismo biológico" (la necesidad de cooperación) es tan esencial para los seres humanos que "la naturaleza nos ha llenado de ideas sobre la necesidad de cooperar":

> Pensamos que debemos ayudar, que tenemos obligaciones para con los demás, porque tener estas ideas va en nuestro interés biológico. Pero desde una perspectiva evolutiva estas ideas existen sencillamente porque aquellos de

> nuestros antepasados que las tuvieron sobrevivieron y se reprodujeron mejor que los que no. En otras palabras, el altruismo es una adaptación humana, igual que lo son nuestras manos y ojos y dientes y brazos y pies. Somos morales porque nuestros genes, modelados por la selección natural, nos llenan de ideas sobre la conveniencia de serlo (RUSE, 2004).

E. Normativismo: ¿es posible construir una ética basada en el funcionamiento de nuestro cerebro? Esto es lo que parecen pensar algunos filósofos y neuroéticos. Patricia Churchland, por ejemplo, critica la idea de que la ciencia no puede decirnos cómo debemos vivir y afirma que, al igual que la salud "es un ámbito en el que la ciencia puede enseñarnos, y ya lo ha hecho, gran parte de lo que deberíamos hacer", también en "el ámbito de la conducta social [...] podemos aprender mucho de la observación común y de la ciencia acerca de las condiciones que favorecen la armonía y la estabilidad social, así como la calidad de vida individual" (CHURCHLAND, 2012). También Gazzaniga propone la construcción de una ética universal basada en el funcionamiento del cerebro (GAZZANIGA, 2015). La idea, tal como la expresa (críticamente) Adela Cortina, sería que "entre el mundo del ser natural y el del deber ser (los códigos morales) existiría un lazo adaptativo que prescribiría establecer como normas éticas aquellas conductas capaces de favorecer la supervivencia" (CORTINA, 2010: 137).

3.2. Cada una de las anteriores afirmaciones, sin embargo, debe verse como una hipótesis, no como una tesis bien establecida, ya que a cada una de ellas se le pueden oponer importantes objeciones aún no respondidas.

La primera tesis, de acuerdo con la cual los juicios morales dependen más de la intuición que de la razón, parece establecer una oposición intuición/razón demasiado radical, esto es, parece estar presuponiendo que una excluye a la otra. Sin embargo, como Bunge ha puesto de manifiesto (BUNGE, 2013: cap. III.1), la idea de "intuición" engloba muchos fenómenos distintos (modos de percepción, formas de imaginación, inferencias rápidas, a saltos o incompletas, capacidad de síntesis, capacidad de evaluación de una situación y de elección de las mejores alternativas...) y varias de esas formas de intuición no pueden verse como opuestas o excluyentes de la razón: la inferencia rápida o incompleta es un razonamiento embrionario o primitivo y la "aprehensión sinóptica –como señala Bunge– no es un

sustituto del análisis, sino un premio al análisis esmerado" (BUNGE, 2013: cap. III.1). Pero, además, los diferentes tipos de intuición –incluso aquellos que no pueden verse en sentido estricto como razonamientos, ni siquiera incompletos– vienen favorecidos especialmente por el ejercicio continuado del razonamiento, del análisis sobre un problema, de la experiencia en una actividad o de la dedicación al estudio de una disciplina. También podría verse la intuición, como ha propuesto Peter Gärdenfors (GÄRDENFORS, 2005), como un tipo de conocimiento implícito y especialmente difícil de explicitar, por no estar estructurado lingüísticamente. En definitiva, de acuerdo con este argumento la contraposición entre intuición y razón sería falsa; la intuición solo se contrapone al razonamiento explícito o proposicional. Las intuiciones a las que se refiere la neuroética podrían ser de este tipo, y la dificultad de dar razones que las justifiquen podría derivarse de la dificultad de acceder a ese conocimiento implícito y no proposicional, pero no de la ausencia de tales razones en la mente del sujeto.

Por lo que respecta a la relación entre el juicio moral y las emociones, también se presentan algunas dificultades. El argumento que está detrás de esta vinculación es el siguiente: por una parte, las imágenes cerebrales demuestran que cuando razonamos moralmente se activan intensamente zonas del cerebro relacionadas con la emoción; por otra, los tests psicológicos demuestran que en la mayoría de los casos resolvemos los dilemas morales de manera intuitiva. Ambas cosas deben estar relacionadas: las intuiciones, por tanto, surgen de las emociones, de la actividad de las zonas afectivas del cerebro. Pero esta manera de reconstruir las decisiones morales plantea, en mi opinión, varios problemas:

a) En primer lugar, la evidencia que tenemos del papel de las emociones en la toma de decisiones es, en realidad, indirecta. Lo que los neurocientíficos pueden comprobar es que cuando nos enfrentamos a un dilema moral las zonas del cerebro relacionadas con las emociones se activan de una manera especialmente acusada. Pero esto no permite inferir que las emociones generan el juicio moral. Podría ser que fuera el juicio moral el que genera la emoción: por ejemplo, darse cuenta, tras un análisis, de lo injusto de una situación puede generar indignación; o, en un dilema moral, ser consciente de que cualquier solución causará un daño puede provocar pesar. De acuerdo con Aristóteles las emociones se relacionan con las creencias en una doble dirección: por un lado, muchas emociones son

generadas por creencias; por otro, muchas emociones generan o modifican nuestras creencias (González Lagier, 2009: 26 y ss.). La neuroética parece estar viendo solo esta segunda conexión. Además, parece haber evidencia de que las emociones se relacionan más con la motivación de la acción moral que con el juicio moral (parece que los individuos psicópatas, a pesar de sus déficits emocionales, tienen intuiciones morales semejantes a los individuos "normales", pero no se ven motivados por ellas).

b) En segundo lugar, la noción de emoción como "marcador somático" las convierte en "impulsos ciegos", meras sensaciones que carecen de contenido proposicional, por lo que solo es posible dar una explicación causal de su relación con los juicios morales (y no una explicación teleológica o basada en razones). Esta manera de entender las emociones y su relación con la moral se aleja de las concepciones de la emoción más extendidas hoy en día entre los filósofos y muchos psicólogos. Para estas concepciones (que, en buena medida, reivindican la concepción de las emociones de Aristóteles), es preciso distinguir en ellas al menos tres dimensiones distintas: 1) una dimensión cognitiva (en un sentido amplio, que incluye desde una creencia hasta una mera percepción); 2) una dimensión afectiva o puramente fenomenológica (la sensación de placer o dolor) y 3) una dimensión motivacional (una tendencia a la acción). Las teorías cognitivas de la emoción centran su atención en el primer aspecto, las teorías "somáticas" y mecanicistas de la emoción se centran en el segundo y las teorías conductistas en el tercero. Por último, las teorías no reductivistas tratan de dar cuenta de todos los aspectos de las emociones. Si se identifican las emociones con el segundo aspecto quedan muchos problemas por resolver: a) no se logra explicar la posibilidad de emociones inconscientes o sin sensación; b) no se logra explicar que las emociones puedan formar parte de explicaciones racionales (teleológicas), y no solo causales, de la conducta; c) no se logra explicar la relación entre las emociones y las creencias; d) no se logra dar cuenta de la posibilidad de evaluar a las emociones como razonables o no (en función de la creencia subyacente), y e) sobre todo, se incurre –de nuevo– en un distanciamiento tajante entre las intuiciones de las que se dice que proceden nuestros juicios morales y la razón (González Lagier, 2009: cap. II).

Tampoco está libre de objeciones la tesis –estrechamente vinculada a las anteriores– del innatismo de las opiniones morales. Los argumentos

a favor del innatismo descansan, una vez más, en el carácter automático de la respuesta y en la incapacidad de dar razones (Haidt), así como en la coincidencia de respuestas a pesar de la heterogeneidad de los encuestados (Hauser). Sin embargo, para asegurar que nuestras creencias morales son innatas habría que descartar completamente que el automatismo se deba a la aceptación de principios profundamente arraigados, pero transmitidos culturalmente. Como señala Adela Cortina comentando los experimentos de Haidt:

> El hecho de que las personas encuestadas respondan de forma intuitiva, es decir, inmediata, automática, sin tener conciencia de cómo han llegado a formular el juicio, y que en muchas ocasiones no sepan dar razón de por qué una acción les parezca buena o mala, puede muy bien explicarse porque lo han aprendido socialmente y no lo han sometido a revisión (Cortina, 2011: 86).

Muchas normas de la moral social (como el incesto, en el ejemplo de Haidt, si es que se acepta su carácter moral) son heredadas del entorno social, que las inculca la mayoría de las veces sin dar razones que las justifiquen. Por otra parte, la universalidad de las normas o principios también debe tomarse con precaución: muchos de los supuestos comportamientos morales universales no son, en realidad, de carácter moral (o exclusivamente de carácter moral): por ejemplo, como hemos visto, Hauser ha postulado la universalidad del principio de cuidado de las crías (que, obviamente, tiene una clara explicación evolutiva), pero ¿deberíamos extraer también la conclusión de que la búsqueda de alimento o el huir de los animales predadores tiene también valor moral? (Bartra, 2013: cap. iii) ¿O de que lo tiene la reproducción? Lo que quiero decir es que muchos hábitos aportados como ejemplos de conductas universalmente aceptadas no poseen (o no poseen exclusivamente) carácter moral.

Hemos visto que otras de las tesis características de este intento de dar cuenta de la moral desde un punto de vista biológico y neurocientífico es su alianza con el darwinismo moral. Las intuiciones morales se ven como mecanismos que la evolución ha seleccionado porque aseguran la supervivencia de la especie. Esta tesis, sin embargo, adolece de cierta ambigüedad. Para aclararla conviene distinguir entre la pretensión (descriptiva) de explicar la moral y la pretensión (de carácter normativo) de

justificarla. Las tesis explicativas, a su vez, pueden tratar de explicar la capacidad del ser humano para tener una conducta ética (para evaluar las conductas como correctas o incorrectas desde el punto de vista moral) o tratar de explicar (lo que es más ambicioso) el contenido de la moral, esto es, tratar de explicar por qué creemos que ciertas conductas son correctas o por qué algunos principios o valores están tan extendidos (Ayala, 2013: 61). Para las tesis explicativas la neuroética recurre a la idea de que comportarse moralmente o adecuar el comportamiento a ciertos principios es un rasgo que ha facilitado la evolución de la especie y su supervivencia. Las tesis normativas añaden que, puesto que eso es así, tales principios están justificados.

De las tesis explicativas creo que debe afirmarse que se trata de hipótesis no suficientemente establecidas. Respecto de la tesis normativa, creo que es directamente el resultado de varios errores.

Tomemos la tesis según la cual lo que explica la capacidad humana de evaluar conductas como buenas o malas y de ajustar la conducta a determinados principios es que esta capacidad es una ventaja evolutiva. Para aceptarla concluyentemente esta tesis debe rechazar la hipótesis alternativa (igualmente plausible) planteada por Francisco Ayala según la cual el comportamiento ético no es directamente un resultado de la evolución, sino que lo es solo indirectamente y en la medida en que es una consecuencia del desarrollo de la inteligencia humana; es decir, lo que tiene valor adaptativo y ha sido favorecido por la evolución es la inteligencia humana, no el hecho de ser capaz de comportarse moralmente (que es consecuencia a su vez de la inteligencia humana) (Ayala, 2013: 66). Si, por el contrario, lo que se afirma es que los códigos morales vienen determinados por la evolución, el problema es que no parece encontrarse un conjunto de principios relevantes, que no estén formulados de una manera excesivamente vaga y vacía, que sean realmente universales; y, además, es posible encontrar tipos de conducta, como la agresividad o la territorialidad, que son importantes evolutivamente y no pueden ser aceptados como ejemplos de conducta moral. La moral que se deriva de la evolución podría ser terrible.

Por su parte, la tesis que trata de reducir lo que debemos hacer a aquello que es bueno para la supervivencia de la especie se enfrenta a la obvia objeción de que viola la ley de Hume: pasa de descripciones acerca de lo que es evolutivamente útil a prescripciones acerca de qué debemos hacer

o cómo debemos vivir. Por ello, los autores que defienden esta postura tratan de argumentar contra la validez de la ley de Hume. Examinemos con más detalle estos argumentos.

3.3. Me parece que los argumentos que usualmente se esgrimen contra la ley de Hume suelen ser de dos tipos:

1) El primer tipo es el de los argumentos basados en contraejemplos: una manera frecuente de mostrar que es posible fundar normas o valores en descripciones consiste en presentar ejemplos de argumentos en los que aparentemente se realiza esta derivación. Esta es la estrategia seguida, en un famoso artículo, por John Searle (SEARLE, 1980: 178-201). Entre los neuroéticos también se ha recurrido a este tipo de argumentos. Marc Hauser, por ejemplo, propone el siguiente:

> *Hecho*: la única diferencia entre un médico que aplica anestesia a una criatura y otro que no se la aplica es que, sin ella, el niño sufrirá enormemente en el curso de una operación. La anestesia no tendrá ningún efecto pernicioso para el niño, sino que le hará perder temporalmente la conciencia y la sensación de dolor. Luego despertará, una vez acabada la operación, sin ninguna secuela negativa y con mejor salud, gracias al trabajo del médico.
>
> *Juicio valorativo*: por consiguiente, el médico debe administrar anestesia al niño (HAUSER, 2008: 28).

Sin embargo, este tipo de argumentos parecen incurrir en uno de los siguientes errores: o bien confunden lo que es bueno o debido desde un punto de vista técnico con lo que es bueno o debido desde un punto de vista moral o normativo, o bien presentan como un argumento completo lo que en realidad es un argumento entimemático que incluye una premisa oculta, que es precisamente la norma o la valoración de la que se deriva la conclusión.

Para evitar el primer error hay que advertir que no siempre que un enunciado incluye el término "deber" es un genuino enunciado normativo. A veces, "debe" expresa una conjetura ("debe ser así" puede significar "probablemente es así"); otras veces, se puede sustituir por "tiene que" y expresa una necesidad práctica. Es importante distinguir entre deberes

deónticos o genuinos y deberes técnicos, prudenciales o necesidades prácticas. Muchos de los ejemplos que se ofrecen como derivación de "debe" a partir de "es" no concluyen genuinos deberes deónticos, sino necesidades prácticas. Como señala von Wright,

> … [podemos] encontrar dos respuestas principales a la cuestión de por qué una cierta cosa debe o puede o no tiene que ser hecha. Una es que existe una norma ordenando o permitiendo o prohibiendo la realización de esa cosa. La otra es decir que los fines y las conexiones necesarias hacen (o no) la realización u omisión de esa cosa una necesidad práctica (VON WRIGHT, 1963: 74).

Por lo que respecta al segundo error, es fácil darse cuenta de que muchas veces los argumentos que formulamos en contextos cotidianos no incluyen todas sus premisas. Es, incluso, factible pensar que en algunos casos es imposible en la práctica enunciar todas las premisas necesarias para llegar a la conclusión. Pero de ello lo que se sigue es la derrotabilidad o revisabilidad de la conclusión, no que su corrección no dependa de la premisa implícita. El ejemplo de Hauser presupone, para su corrección, una premisa según la cual se debe evitar el sufrimiento innecesario.

(2) El segundo tipo de argumentos restringe la ley de Hume a los argumentos deductivos: se afirma que lo que la ley de Hume proscribe es derivar deductivamente un deber ser a partir del ser, pero que existen otro tipo de inferencias aceptables, como la inducción o la inferencia a la mejor explicación, por medio de las cuales sí se puede pasar de descripciones de hechos a normas (CASEBEER, 2003: 482; CHURCHLAND, 2012: cap. 1). Una manera de defender que la ley de Hume se refiere exclusivamente a inferencias deductivas consiste en entenderla como una consecuencia del principio de conservación de la lógica: en una deducción no es posible que se concluya algo que no estuviera incluido ya en las premisas. Las deducciones pueden hacer que seamos conscientes de un dato nuevo, pero este se encontraba ya en las premisas. Por ello, solo de proposiciones descriptivas no podemos deducir enunciados de deber ser. Esto ocurre con cualquier cosa. Como observa Pigden, "existe un salto similar entre las conclusiones sobre los erizos y las premisas que no hacen mención de ellos. No se pueden obtener conclusiones sobre 'erizos' a partir de premisas carentes de erizos (al menos no solo por la lógica)" (PIGDEN, 2004: 570). Ahora

bien, a diferencia de las deducciones, las inducciones y las abducciones sí amplían nuestro conocimiento, por lo que no rige para ellas el principio de conservación. Por tanto, si la ley de Hume es solo una manifestación del principio de conservación de los argumentos deductivos, entonces tienen razón los que sostienen que esta no es aplicable a las inferencias no deductivas.

¿Pero es solo eso la ley de Hume? Probablemente, no. Puede sostenerse que entre enunciados descriptivos y enunciados normativos existen importantes diferencias –a veces se habla de un "abismo lógico" entre ellos, o entre hechos, por un lado, y normas y valores, por otro–: así, los enunciados descriptivos tienen una *dirección de ajuste*[8] descendente (esto es, palabras-a-mundo: se pretende que las palabras se ajusten al mundo) mientras que los enunciados normativos tienen una dirección de ajuste ascendente (esto es, mundo-a-palabras: se pretende que el mundo se ajuste a las palabras). Los enunciados descriptivos son verdaderos o falsos, mientras que las normas o valores no lo son. Los enunciados que expresan deberes presuponen un punto de vista interno, en el que los términos deónticos (obligatorio, prohibido) son usados, mientras que si una descripción se refiere a una norma o a un deber, lo hace desde un punto de vista externo en el que los términos deónticos son solo mencionados. Todas estas diferencias entre descripciones y normas hacen que las primeras no puedan servir de razones, ni expresar razones, para justificar las segundas. No es solo que una justificación deductiva requiera que entre las premisas esté aquello que se quiere deducir, es que –aun cuando se admita que la inducción o la abducción puedan tener alcance justificatorio– ningún enunciado descriptivo puede, por sí solo, ser una razón que justifique un enunciado prescriptivo. Puede aportar una razón explicativa de por qué aceptamos ciertas normas o valores; puede ser también una razón explicativa o, incluso, justificatoria de otros enunciados descriptivos. Pero si se quiere concluir la justificación de una norma a partir de descripciones y por medio de argumentos no deductivos, se tiene la carga de la prueba. Y es interesante observar que los neuroéticos no lo han hecho.

---

8 Una explicación de la idea de "dirección de ajuste" puede encontrarse en Searle, 1976.

3.4. Si las anteriores consideraciones son correctas, cuando se propone que debemos seguir aquellas pautas de conducta que tienen un valor adaptativo, o bien simplemente estas se recomiendan como medidas prudenciales para mantener la supervivencia de la especie humana, pero entonces no tienen carácter moral, o bien se asume que la supervivencia de la especie es un fin moralmente valioso, en cuyo caso la normatividad no viene de los hechos, sino de esta asunción valorativa. ¿Quiere decir todo lo anterior que la neurociencia no puede aportar nada relevante para la comprensión de la normatividad? Esta sería una conclusión equivocada. Todo razonamiento práctico-normativo (esto es, si no se trata de necesidades prácticas o deberes técnicos) tiene una premisa normativa y una premisa fáctica. La neurociencia puede contribuir al establecimiento de esta premisa fáctica y, más aún, si asumimos el principio "debe implica puede", puede establecer límites a las normas que tiene sentido establecer: esto es, la neuroética podría ayudarnos a entender cuál es el espacio de la moralidad. Pero no puede justificar por sí sola las respuestas a los problemas éticos.

4. ALGUNOS PROBLEMAS DE LA NATURALIZACIÓN DE LA MENTE

4.1. El intento de naturalizar la mente plantea también importantes objeciones filosóficas. Aquí me referiré a dos de ellas, que podemos llamar la "objeción eliminacionista" y la "objeción de la prioridad epistémica" de lo mental sobre lo neuronal" (señalada por G. H. von Wright).

De acuerdo con el primer argumento, las propuestas más radicales de naturalización de la mente basadas en las aportaciones de la neurociencia identifican dos fenómenos que, en realidad, son distintos: la mente y el cerebro. Por ejemplo, para los defensores de la tesis de la identidad entre estados mentales y estados cerebrales todos los conceptos mentales pueden (o podrán) ser traducidos al lenguaje de los estados cerebrales sin ninguna pérdida significativa y, por tanto, la psicología puede traducirse por completo al lenguaje de la neurociencia. A lo sumo, el reductivismo acepta que puede ser útil que sigamos hablando de deseos, creencias, etc. y mantengamos una "psicología popular" (el eliminacionismo ni siquiera admite esto), pero solo como un lenguaje no científico y siendo conscientes de que hace referencia a entidades inexistentes (MOYA, 2006; CHURCHLAND, 1999).

Muchos autores han señalado que esta es una postura excesivamente radical[9]. Los estados mentales tienen ciertos rasgos (en lo que sigue me referiré a ellos como "las propiedades de lo mental") que parecen encajar mal en una concepción materialista y naturalizada del ser humano. En primer lugar, nuestros estados mentales nos son accesibles a nosotros mismos, por introspección, de una manera directa, al margen de la evidencia empírica y de inferencias a partir de ella (es el rasgo de la conciencia). En segundo lugar, y relacionado con lo anterior, hay cierta diferencia cualitativa en la manera como emergen en mi conciencia los dolores, deseos o preocupaciones, o en la manera como experimento escuchar un concierto de Bach o el sabor de un trozo de chocolate. Puede decirse, entonces, que estos estados mentales se corresponden con diferentes sensaciones internas, que los filósofos –en analogía con el término *quanto* usado en física– llaman *qualia*. En tercer lugar, muchos estados mentales (las creencias, las intenciones, etc.) poseen un contenido, un significado, versan sobre otros hechos (es el rasgo de la "intencionalidad" o "contenido mental"). En cuarto lugar, los estados mentales se relacionan con nuestra conducta externa, pareciendo tener eficacia causal sobre ella, pero no en virtud de su dimensión física, sino de su contenido representacional (es el fenómeno de la causación mental). Consciencia, carácter cualitativo, contenido representacional y eficacia causal son cuatro características de la mente que no es evidente que puedan explicarse por referencia exclusiva a procesos físico-químicos y leyes empíricas. Supongamos que la mente se identifica con el cerebro y que los estados mentales son conexiones eléctricas, físicas o químicas entre neuronas: ¿cómo es posible que algo material, físico, nos resulte accesible sin observación externa y posea un contenido semántico?[10]. Y si la mente no forma parte del mundo físico, ¿cómo puede interactuar causalmente con él?

La alternativa parece ser, entonces, mantener el dualismo cartesiano mente-cerebro o asumir el materialismo reductivista, con lo que habría que llevar los estados mentales al terreno de la normatividad: se verían como

---

9 En relación con el derecho, véanse Taruffo, 2013, y Pardo y Patterson, 2011.

10 Una presentación de una estrategia para la naturalización de la intencionalidad puede verse en Moya, 1994.

ficciones normativas que deben presumirse en ciertas circunstancias, sin importar su correspondencia con estados subjetivos de los agentes, pues tales estados en realidad no existen como entidades mentales. Sin embargo, hay una salida a este dilema, porque el materialismo no tiene por qué ser eliminacionista o reductivista. Un tipo de materialismo no reductivista es el emergentismo de autores como John Searle o Mario Bunge (Bunge, 2002).

El emergentismo es una postura materialista que identifica los estados mentales con propiedades emergentes o sistémicas, es decir, propiedades que surgen a partir de cierto grado de complejidad de un sistema. Son propiedades del conjunto del sistema, pero no de cada una de sus partes. Para explicar qué es una propiedad emergente John Searle recurre al siguiente ejemplo:

> Supongamos que tenemos un sistema S, compuesto de los elementos a, b, c... Por ejemplo, S podría ser una piedra y los elementos podrían ser moléculas. En general, habrá rasgos de S que no son, o no necesariamente, rasgos de a, b, c... Por ejemplo, S podría pesar 20 kilogramos, sin que las moléculas individualmente pesen 20 kilogramos. Denominemos a estos rasgos 'rasgos del sistema'. Algunos rasgos del sistema pueden ser deducidos, o determinados, o calculados a partir de los rasgos a, b, c..., simplemente por la forma en que se componen y ordenan (y, a veces, por sus relaciones con el entorno). Ejemplos de ellos serían la forma, el peso y la velocidad. Pero algunos otros rasgos del sistema no pueden ser determinados solo a partir de los elementos que los componen y de las relaciones con el entorno: han de ser explicados a partir de las relaciones entre los elementos. Llamémosles 'rasgos del sistema causalmente emergentes'. La solidez, la liquidez y la transparencia son ejemplos de rasgos del sistema causalmente emergentes (Searle, 1996: 121).

De manera que, para Searle, una propiedad emergente es una propiedad que surge a partir de los elementos que componen el sistema, de las relaciones del sistema con el entorno y –esto parece lo decisivo– de las relaciones de esos elementos entre sí. Los estados mentales se identifican con propiedades emergentes de los cerebros. Los cerebros son sistemas físicos que han ido adquiriendo complejidad a lo largo de su historia evolutiva. El naturalismo biológico es descrito por Searle a partir de cuatro tesis: 1) Los estados mentales, con su ontología subjetiva de primera persona, son fenómenos reales del mundo real; 2) Los estados mentales son causados

en su totalidad por procesos neurobiológicos de nivel inferior localizados en el cerebro; 3) Los estados mentales son rasgos del sistema cerebral en su conjunto y existen, por tanto, en un nivel superior al de las neuronas y sus sinapsis ("Por sí misma –dice Searle–, una neurona no es consciente, pero las partes del sistema cerebral compuestas por ella sí lo son"), y 4) Los estados mentales, en la medida en que son reales, tienen eficacia causal ("Mi sed consciente, por ejemplo, me lleva a tomar agua") (SEARLE, 2006: 147-148).

El materialismo emergentista no es reductivista. De acuerdo con Searle, aunque los estados mentales pueden ser reducidos causalmente a –en el sentido de explicados por– los componentes del cerebro y sus relaciones, esta explicación no es suficiente, porque lo característico de lo mental es que es un fenómeno en primera persona, requiere un lenguaje subjetivo que no puede ser descrito con el lenguaje objetivo de la neurociencia. El emergentismo es un monismo en cuanto a la sustancia, pero mantiene un dualismo de propiedades, siendo las propiedades mentales irreductibles a propiedades físicas o neuronales. La neurociencia, si esta es una visión correcta de la relación entre lo cerebral y lo mental, no puede reemplazar a la psicología, pero es necesaria para explicar la relación causal entre lo neuronal y lo mental.

4.2. El segundo argumento contra la naturalización reductivista de la mente a la neurociencia es el que he llamado "prioridad epistémica de lo mental sobre lo neuronal", señalada por G. H. von Wright.

Como hemos visto, lo que los neurocientíficos parecen decir muchas veces es que hay correlatos empíricos (neuronales) de los estados mentales y que una vez conocida esta correlación la conducta humana se podrá explicar completamente en términos del sustrato neuronal. Lo que von Wright pretende, por el contrario, es mostrar que no se pueden eliminar los estados mentales de las explicaciones de la conducta[11].

---

11 Sigo su argumento tal como fue formulado en la lección de recepción del doctorado *honoris causa* en la Universidad de Leipzig el 21 de mayo de 1996, con el título *La posición de la psicología entre las ciencias*. Agradezco a Ernesto Garzón Valdés y a Francisco Laporta el haberme llamado la atención sobre esta conferencia y el haberme

De acuerdo con von Wright el estudio de los fenómenos mentales, esto es, la psicología científica, puede realizarse desde tres enfoques distintos: el primer enfoque consiste en analizar lo mental desde el punto de vista de la introspección o "auto-observación". Los estados mentales se caracterizan porque no son intersubjetivamente observables y, en ese sentido, no son objetivos; son "propiedad privada del sujeto que los tiene", que dispone de un acceso directo a los mismos (sabemos, sin necesidad de observarnos "desde fuera", qué intenciones tenemos, qué creencias sostenemos, qué emociones experimentamos, etc.). A este enfoque lo podemos llamar la "psicología de la consciencia" (la cual a veces se ha denominado también, en un sentido despectivo, "psicología popular"). Pero lo mental tiene un sustrato físico o corporal, que a su vez presenta otros dos aspectos: lo mental, por un lado, se expresa en la conducta, en los movimientos corporales del sujeto y los cambios que producen en el mundo; y, al mismo tiempo, también está constituido por sucesos intracorporales que el sujeto no puede, en principio, observar en sí mismo: los fenómenos neuronales que ocurren en el cerebro del sujeto. Al primer aspecto, von Wright lo llama el aspecto conductual de lo mental; al segundo, el aspecto neuronal. La psicología que se desarrolla estudiando el primer aspecto es la psicología de la conducta o conductista; la que se desarrolla analizando el segundo aspecto es la neuropsicología.

Como señala von Wright, las tres ramas de la psicología "no se llevan bien" y la psicología de la conducta y la neuropsicología sospechan de la psicología de la consciencia, hasta el punto de que han pretendido eliminarla. El conductivismo clásico de John B. Watson intentó reducir la psicología de la consciencia a la conductista, sosteniendo que la consciencia y los estados mentales no existen y son solo fenómenos conductuales complejos. La neuropsicología o neurociencia (al menos sus defensores más radicales), por su parte, pretende reducir los estados mentales a los fenómenos neuronales. Esta reducción no es posible, según von Wright,

---

proporcionado el texto. Las citas entrecomilladas, salvo que se indique otra cosa, pertenecen al texto de esa conferencia. También se apuntan algunos aspectos de este argumento de von Wright (aunque menos definidos) en VON WRIGHT, 2002, y en VON WRIGHT, 2003.

dadas las relaciones peculiares que lo mental, lo conductual y lo neuronal mantienen entre sí. Su argumento consiste en mostrar el juego de relaciones y prioridades que se dan entre estas tres dimensiones de los estados de consciencia. Estas relaciones son la *prioridad causal* de lo neuronal frente a lo conductual, la *prioridad epistemológica* de lo mental sobre lo neuronal y la *prioridad semántica* de lo conductual frente a lo mental. Veámoslas.

Supongamos que se produce de pronto un ruido y el sujeto A vuelve la cabeza inmediatamente hacia el lugar de donde este procede. Podemos explicar el movimiento corporal aludiendo a ciertos procesos fisiológicos. En palabras de von Wright,

> … las ondas sonoras se introducen en el oído interno y provocan allí procesos sensoriales centrípetos que se propagan hacia el centro de la audición en el cerebro. Desde allí son transmitidos de nuevo a un centro motor y originan impulsos motores centrífugos que se propagan hasta los músculos y, finalmente, dan como resultado movimientos del cuerpo.

Esta es una explicación fisiológica del movimiento corporal como reacción a un estímulo. Si queremos explicar causalmente cómo se produjo el movimiento corporal debemos recurrir a este tipo de explicaciones, en las que lo neuronal aparece como causa de la conducta (del movimiento corporal). En esto consiste la *prioridad causal* de lo neuronal sobre lo conductual.

Supongamos ahora que le preguntamos al sujeto por qué ha vuelto la cabeza y este responde que lo ha hecho porque ha oído un ruido y, dado que estaba esperando a alguien, quería saber si ya había llegado. Ahora tenemos, junto con la explicación fisiológica, una explicación racional de la conducta. La explicación neuronal y la explicación racional se sitúan en niveles distintos, pero en cierto sentido hay una correlación entre ellos: debe haber un sustrato neuronal de "percibir un sonido" y un sustrato neuronal de la razón aducida para volver la cabeza ("querer averiguar si es la persona que esperaba"). De manera que sería posible, a partir de observaciones del sistema nervioso, descubrir si una persona ha oído un sonido o, incluso, si tiene cierto deseo, o una creencia, etc. Por ejemplo, cuando detectáramos en el sujeto A la actividad cerebral x podríamos decir que el sujeto ha oído un ruido. Ahora bien, para llegar a esto previamente hemos tenido que establecer una correspondencia entre "oír ruidos" y la

actividad cerebral x, y para establecer inicialmente esta correspondencia necesitamos criterios distintos de los neuronales para identificar que el sujeto está oyendo algo. Y lo mismo ocurre con el correlato neuronal de las razones para hacer u omitir algo. En palabras de von Wright:

> El hecho, por ejemplo, de que ciertas alteraciones hormonales sean indicativas de un estado de miedo o cansancio es algo que se ha podido establecer sobre la base de investigaciones anatómico-fisiológicas en seres vivos de los que ya se sabía que estaban asustados o cansados. Y para saber esto debemos saber ya qué significa estar asustado o cansado y saber cómo se puede comprobar eso sin tener que apelar a criterios intracorporales.

Lo mental tiene, por tanto, *prioridad epistemológica* sobre lo neuronal: para establecer cuál es el correlato neuronal de estados mentales como "oír un sonido", "tener miedo", "estar sediento", etc. necesitamos previamente tener ya identificados estos estados mentales. Solo si ya los tenemos identificados, podemos descubrir cuáles son sus correlatos neuronales. Por el contrario, observando solo los continuos procesos neuronales no podremos saber a qué estado mental se corresponden.

Surge ahora una nueva pregunta: ¿qué criterios usamos para identificar lo mental? La respuesta de von Wright, siguiendo a Wittgenstein, es que descubrimos que un sujeto tiene una u otra razón o está en uno u otro estado mental a través de su conducta externa.

> ¿Cómo sabemos si un animal, por ejemplo un perro, ha oído un sonido? Normalmente porque vuelve la cabeza en dirección al sonido o de otro modo porque adopta una actitud de atención, o echa a correr si está asustado, o corre por el contrario en dirección al sonido si quiere averiguar lo que pasa.

En el caso de los seres humanos, la conducta que usamos como criterio es frecuentemente –pero no siempre– verbal. Cuando le preguntamos al sujeto de nuestro ejemplo por qué ha vuelto la cabeza y nos responde que porque quería comprobar si ya había llegado la persona que esperaba su declaración es un tipo de conducta –conducta verbal– que nos indica qué razón tenía. El resto de su comportamiento (no verbal) nos ayuda a confirmar que esta es realmente su razón. Usando la distinción wittgensteiniana entre síntoma y criterio podríamos decir que lo neuronal es síntoma de lo

mental, pero lo conductual es algo más fuerte: es el criterio que usamos para determinar que un sujeto está bajo uno u otro estado mental (VON WRIGHT, 2002). Von Wright sugiere que la relación entre la conducta y lo mental debe entenderse como una relación semántica: "Aquéllos [los criterios conductuales] nos dicen qué quiere decir o significa que, por ejemplo, un sujeto oiga un ruido o esté asustado por algo o esté cansado". En esto consiste la *prioridad semántica* de lo comportamental frente a lo psíquico o mental. Sin conducta externa no podríamos entender qué quiere decir estar en uno u otro estado mental, ni identificar que un sujeto está bajo uno u otro estado mental.

Lo neuronal, por tanto, causa la conducta; lo mental es necesario para poder identificar lo neuronal (y dota de sentido a la conducta) y lo conductual es el criterio que usamos para comprobar lo mental. Dadas estas tres prioridades, ninguna de estas dimensiones puede reducirse a otra. Lo conductual nos es necesario para entender los estados mentales de los demás (y los nuestros) y como criterio de identificación de los mismos, y tener identificados los estados mentales es necesario para encontrar sus correlatos neuronales. Todo intento de reduccionismo conlleva una pérdida importante en nuestra capacidad de comprendernos a nosotros mismos. Cuando la neurociencia pretende reducir lo mental a lo neuronal arguyendo que lo neuronal tiene prioridad causal frente a lo mental se olvida del resto de dimensiones de la relación, respecto de las cuales lo neuronal es *secundario*.

REFLEXIONES FINALES

1. En las anteriores páginas se han revisado algunos de los problemas que plantea la naturalización de la filosofía (práctica). Podría distinguirse entre dos maneras de entender este proceso de naturalización: una manera radical (la tesis del reemplazo), que trata de disolver la filosofía en la ciencia empírica (actualmente, en la neurociencia), y otra, más moderada (la tesis de la complementariedad), que recuerda la necesidad de que la filosofía no contradiga la visión del mundo (y de la naturaleza del hombre) que está surgiendo de las ciencias, pero que no niega que la filosofía tiene un espacio donde la ciencia no puede llegar. Incluso, desde este segundo punto de vista, puede afirmarse que la filosofía tiene una tarea (el esclarecimiento

conceptual) sin la cual la ciencia no puede proporcionar conclusiones relevantes acerca de las normas, la mente, la acción, etc. Las aportaciones de las ciencias han de ser interpretadas, y en esa interpretación la filosofía ejerce una labor de control.

2. En el actual estado de desarrollo de la neurociencia, el programa de naturalización de la tesis del reemplazo no puede ser realizado: nuestro conocimiento del funcionamiento del cerebro no permite naturalizar la normatividad sin incurrir en la falacia naturalista o confundir normas y regularidades, ni naturalizar la mente sin eliminar las propiedades de lo mental o ignorar la complejidad de las relaciones entre lo neuronal, lo mental y la conducta.

3. No obstante lo anterior, la filosofía debe tener en cuenta las aportaciones de la ciencia: la neurociencia puede ayudarnos a comprender cuál es el ámbito en el que tienen sentido las exigencias morales (por ejemplo, ayudándonos a entender las fronteras de nuestro libre albedrío) y cómo surgen las propiedades mentales a partir de la actividad neuronal. Podría distinguirse entre un "monismo ontológico" (la aceptación de que solo existe un tipo de entidades: las físicas) y un "monismo metodológico" (la tesis de que solo es legítimo el método de las ciencias empíricas). El estado actual de la ciencia parece avalar el primero, pero no es suficiente para aceptar el segundo. Todo ello sugiere que la tesis de la complementariedad es un proyecto de naturalización más viable y razonable que la del reemplazo.

REFERENCIAS

Ayala, F., 2013: *Evolución, ética y religión*. Bilbao: Universidad de Deusto.

Bartra, R., 2013: *Cerebro y libertad. Ensayo sobre la moral, el juego y el determinismo*. México: FCE (versión electrónica).

Bonete Perales, E., 2010: *Neuroética práctica. Una ética desde el cerebro*. Bilbao: Desclée de Brouwer.

Bunge, M., 2002: *El problema mente-cerebro. Un enfoque psicobiológico*. Madrid: Tecnos.

Bunge, M., 2013: *Intuición y razón*. Buenos Aires: Debolsillo (versión electrónica).

Casebeer, W. D., 2003: "Moral cognition and its neural constituents", *Nature Reviews/Neuroscience*, 4.

Chaungeux, J. P., 1985: *El hombre neuronal*. Madrid: Espasa-Calpe.

Churchland, P., 1999: *Materia y conciencia*. Barcelona: Gedisa.

Churchland, P., 2012: *El cerebro moral. Lo que la neurociencia nos cuenta sobre la moralidad*. Barcelona: Paidós (versión electrónica).

Cortina, A., 2010: "Neuroética: ¿las bases cerebrales de una ética universal con relevancia política?", *Isegoría, Revista de Filosofía Moral y Política*, 42.

Cortina, A., 2011: *Neuroética y neuropolítica. Sugerencias para la educación moral*. Madrid: Tecnos.

Damasio, A. R., 2004: *El error de Descartes*. Barcelona: Crítica.

Garderförs, P. 2005: "Las intuiciones como conocimiento incompleto", *Revista de Filosofía*, Universidad de Costa Rica, xliii, 108.

Gazzaniga, M. S., 2015: *El cerebro ético*. Barcelona: Paidós.

Giménez Anaya, J. M. y Murillo, J. I., 2007: "Mente y cerebro en la neurociencia contemporánea. Una aproximación a su estudio interdisciplinar", *Scripta Theologica*, 39.

González Lagier, D., 2009: *Emociones, responsabilidad y derecho*. Madrid: Marcial Pons.

González Lagier, D., 2017: *A la sombra de Hume. Un balance crítico del intento de la neuroética de fundamentar la moral*. Madrid: Marcial Pons.

Greene, J. D., 2012: "Del 'es' neuronal al 'debe' moral: ¿cuáles son las implicaciones morales de la psicología moral neurocientífica?", en Cortina, A. (ed.), *Guía Comares de neurofilosofía práctica*. Granada: Comares.

GRIMALDOS, T. y PACHO, J. (eds.), 2005: *La naturalización de la filosofía: problemas y límites*. Valencia: Pre-Textos.

HAIDT, J., 2012: "El perro emocional y su cola racional: un enfoque intuicionista social del juicio moral", en CORTINA, A. (ed.), *Guía Comares de neurofilosofía práctica*. Granada: Comares.

HAUSER, M. D., 2008: *La mente moral. Cómo la naturaleza ha desarrollado nuestro sentido del bien y del mal*. Barcelona: Paidós.

MARTÍNEZ, S. F. y OLIVÉ, L., 1997: "Introducción", en MARTÍNEZ, S. F. y OLIVÉ, L. (comps.), *Epistemología naturalizada*. Barcelona: Paidós.

MOYA, C., 1994: "Las emociones y la naturalización de la intencionalidad", *Anales del Seminario de Metafísica*, 28.

MOYA, C., 2005: "La naturalización de la responsabilidad moral", en GRIMALDOS, T. y PACHO, J. (eds.), *La naturalización de la filosofía: problemas y límites*. Valencia: Pre-Textos.

MOYA, C., 2006: *Filosofía de la mente*. Valencia: Universidad de Valencia.

NAVARRO, P., 2007: "Normas condicionales y falacia naturalista", *Doxa, Cuadernos de Filosofía del Derecho*, 30.

PARDO, M. S. y PATTERSON, P., 2011: "Fundamentos filosóficos del derecho y la neurociencia", *InDret*, 2.

PIGDEN, Ch. R., 2004: "El naturalismo", en SINGER, P., *Compendio de ética*. Madrid: Alianza Editorial.

QUINE, W. V. O., 2002: "Naturalización de la epistemología", en *La relatividad ontológica y otros ensayos*. Madrid: Tecnos.

ROSKIES, A., 2002: "Neuroethics for the New Millenium", *Neuron*, 35, 1.

RUSE, M., 2004: "La significación de la evolución", en SINGER, P., *Compendio de ética*. Madrid: Alianza Editorial.

Searle, J., 1976: "Una taxonomía de los actos ilocucionarios", *Teorema*, VI/1.

SEARLE, J., 1980: *Actos de habla.* Madrid: Cátedra.

SEARLE, J., 1996, *El redescubrimiento de la mente.* Barcelona: Crítica.

SEARLE, J., 2006, *La mente. Una breve introducción.* Norma.

TARUFFO, M., 2013, "Proceso y neurociencia", en TARUFFO, M. y NIEVA, J. (dirs.), *Neurociencia y proceso judicial.* Madrid: Marcial Pons.

VON WRIGHT, G. H., 1963, *The Varieties of Goodness.* London: Routledge & Kegan Paul.

VON WRIGHT, G. H., 1996, *La posición de la psicología entre las ciencias*, lección de doctorado "honoris causa" en la Universidad de Leipzig, 21 de mayo de 1996.

VON WRIGHT, G. H., 2002: *Sobre la libertad humana.* Barcelona: Paidós.

VON WRIGHT, G. H., 2003: "Valor, norma y acción en mis escritos filosóficos. Con un epílogo cartesiano", *Doxa*, 26.

II.
Acción, expresión y atribución de estados mentales[**]

SEBASTIÁN FIGUEROA RUBIO[*]

¡Ah, poder expresarme todo yo como un motor
se expresa!
¡Ser completo como lo es una máquina!
Álvaro de Campos, *Oda triunfal*

INTRODUCCIÓN

Es fácil constatar que el uso de conceptos mentales o psicológicos juega un rol relevante en el derecho. No solo en la práctica jurídica en la que se hace referencia a dolo, deliberación y consentimiento, sino también en la filosofía del derecho que suele hablar de aceptación de normas o de decisión judicial. Es inescapable para lograr una comprensión del fenómeno jurídico ser conscientes de lo que el uso de estos conceptos conlleva.

Si bien la reflexión filosófica sobre estos conceptos se puede rastrear en la Antigüedad, es durante la Modernidad que adquiere especial fuerza, siendo parte sustancial de los sistemas filosóficos y del legado de autores como René Descartes, John Locke e Immanuel Kant. No es de extrañar que una de las ramas de la filosofía que mayor desarrollo ha tenido durante el siglo XX sea la denominada filosofía de la mente o filosofía de la psicología. Es esperable, a su vez, que las discusiones filosóficas tengan efectos en el ámbito jurídico, no solo en la práctica del derecho, sino también en la forma en que los filósofos del derecho enfrentan su objeto de reflexión. Teniendo esto en cuenta, en el presente trabajo quisiera revisar críticamente algunos de los supuestos de una perspectiva cada vez más popular sobre estos temas en el ámbito jurídico que tiene que ver con los efectos para la comprensión del uso de conceptos mentales de la introducción de la neurociencia cognitiva.

---

* Doctor en Derecho por la Universitat de Girona (España). Profesor asistente de la Universidad Adolfo Ibáñez (Viña del Mar, Chile).

** El autor agradece los comentarios realizados por Nicola Muffato y Juan Ormeño a una versión previa de este texto. También agradece las conversaciones sobre el asunto con Felipe de la Fuente, Maribel Narváez y Josep Lluís Prades, y a Bernardo Mundaca por su colaboración en la edición del texto.

Este trabajo es un producto del proyecto Fondecyt 11180421 del cual el autor es investigador responsable.

Es posible identificar tres tesis que usualmente se combinan al hablar de los efectos de la neurociencia cognitiva en la comprensión de la mente. Tanto la defensa de dichas tesis como su combinación repercuten en nuestro entendimiento del derecho y, en conjunto, forman argumentos que deben ser revisados críticamente[1]. Al estar presentes en dichos argumentos y por la forma en que se relacionan, vale la pena hablar de pasos que se van dando.

El primero de estos pasos considera a la naturaleza misma de la neurociencia como estudio del cerebro[2] y consiste en comprometerse con que el hecho primario al hablar de lo mental es un grupo de fenómenos que se dan en el cerebro. El segundo paso se entiende como una consecuencia del anterior y supone identificar a los estados mentales de las personas con estados en que se encuentra el cerebro en cuanto órgano del cuerpo. El tercer paso consiste en afirmar que la agencia es un producto de los estados físicos en que se encuentra el cerebro y que, por ende, para comprender *realmente* los aspectos mentales propios de la agencia se debe prestar atención al funcionamiento del cerebro[3]. Este último paso se conforma como una manera de ver la relación entre lo mental y el comportamiento de las personas, y se presenta como la aplicación de lo dicho en el segundo paso al ámbito de lo práctico. Así, por ejemplo, se ha argumentado que una decisión es algo que ocurre primeramente en el cerebro y que nuestra percepción de lo que ocurre en el cerebro nos lleva a elaborar la idea de que optamos entre alternativas. A partir de ello se afirma que una correcta

---

1 Para las fuentes, su desarrollo histórico y los autores más importantes detrás de estas ideas, véase Bennett y Hacker, 2005: parte 1; Freeman, 2011; Narváez, 2016.

2 Hacker y Bennett explican esto de la siguiente forma: "Las preguntas empíricas sobre el sistema nervioso son el campo de la neurociencia. El cometido de esta es establecer los hechos en lo que concierne a las estructuras y las operaciones neurales. Es tarea de la neurociencia cognitiva explicar las condiciones neurales que hacen posibles las funciones perceptivas, cognitivas, cogitativas, afectivas y volitivas. Las investigaciones experimentales confirman o cuestionan tales teorías explicativas" (Bennett y Hacker, 2007: 15; véase Narváez, 2016: 98-107).

3 La noción de agencia es muy amplia, en este trabajo considero aquella que constituye la libertad de acción y que se suele identificar con competencias normativas de diversas clases. Es esta noción de agencia la que se vincula con conceptos como intención, decisión y dolo.

comprensión de lo que es decidir y actuar conforme a nuestras decisiones depende de conocer los mecanismos que ocurren en el cerebro (Libet, 1985). Bajo esta perspectiva, decidir es algo que *el cerebro hace*, por lo que para acceder a un estado mental es necesario acceder a estados cerebrales.

De forma más esquemática, las tesis que están presentes en estos tres pasos pueden identificarse así:

Primero, para hablar de lo mental se debe hablar de lo que ocurre en el cerebro.

Segundo, los estados mentales son estados cerebrales.

Tercero, los elementos mentales que constituyen la agencia se encuentran en el cerebro.

Esta forma de presentar los pasos no hace referencia a las premisas que llevan a adoptar dichas tesis, ni tampoco muestra que entre ellos conformen un razonamiento. Esto es así, en primer lugar, porque hay diversas formas de interpretar los términos utilizados al expresar estas tesis, por lo que habrá que realizar ciertas especificaciones al respecto. En segundo lugar, porque parte de este trabajo consiste en mostrar algunos de los argumentos más influyentes en que se sustenta la adopción de dichas tesis, haciendo explícitas conexiones entre ellas. Esto permitirá ver, a su vez, de qué manera tiene sentido entenderlas como pasos de un camino argumental.

A partir de la adopción de las referidas tesis se pueden identificar dos actitudes filosóficas respecto de los conceptos mentales que han dominado la filosofía de la psicología en las últimas décadas. Según la primera, que podemos denominar reduccionista, debe ser posible traducir cualquier concepto mental en estados cerebrales para que tenga sentido hablar sobre ellos; según la segunda, denominada eliminativista, debemos abandonar el uso de dichos conceptos por ser parte de una teoría fallida. Más allá de sus diferencias, ambas actitudes abrazan un conjunto de teorías y supuestos que permiten identificarlas como parte de la perspectiva estándar en estas materias[4]. En este trabajo quisiera reflexionar acerca de dichos supuestos comunes, aunque para ello me centraré en las ideas reduccionistas debido

---

4 Para una revisión crítica detallada de las diferencias e ideas comunes entre las opciones de esta visión estándar, véase Rudder Baker, 1995: parte I.

a que han sido defendidas con mayor fuerza en el ámbito jurídico[5]. En efecto, sea cual sea la actitud que tomemos, una consecuencia de estos tres pasos es la popularización de eslóganes tales como "la libertad es una ilusión", "yo no actúo, mis neuronas son las que lo hacen" o "debemos abandonar conceptos como responsabilidad y culpabilidad". Estas últimas afirmaciones han sido largamente discutidas y acá no pretendo hacerme cargo de ellas, más bien quisiera argumentar en contra de un camino que supone seguir los tres pasos referidos y que conduce a una forma particular de entender la atribución de estados mentales, teniendo en cuenta lo que implica atribuir estados mentales en el contexto jurídico.

Este texto tiene tres partes. En la primera se hacen explícitas algunas ideas filosóficas en las que se sostienen los tres pasos anunciados (secciones 1 y 2). En la segunda se someten a crítica dichas ideas, en especial considerando sus consecuencias para la comprensión de la relación entre lo mental y el comportamiento de las personas (sección 3). Por último, se presentan lineamientos de inspiración wittgensteiniana con el fin de defender una noción de agencia y de lo mental distinta que, a mi parecer, puede ser de gran utilidad para reflexionar acerca de lo jurídico, en especial cuando hablamos de atribuir estados mentales en dicho contexto (secciones 4 y 5).

1. DUALISMO DE SUSTANCIAS Y FISICALISMO CAUSAL

Antes de comenzar a revisar las ideas filosóficas en que se sostienen los pasos señalados, hay que tener presentes ciertas características de los conceptos mentales o psicológicos que marcan la reflexión en torno a ellos y que nos permitirán delimitar las preguntas que guían lo que se dirá a continuación. En primer lugar, lo mental está compuesto por un conjunto heterogéneo de estados. Entre ellos contamos con actitudes proposicionales (como creencias y deseos), sensaciones perceptuales (como las auditivas y las visuales), corporales (como el hambre y la sed) y emociones (como la

---

5 Las propuestas centrales del eliminativismo se encuentran en los trabajos de Patricia Churchland y Paul Churchland. Una revisión crítica puede encontrarse en Moya, 2006: cap. 7; Ramsey, 2019; Toribio, 1995, y, en el ámbito jurídico, en Pardo y Patterson, 2013: 28-42.

ira y la alegría). Esto conlleva una gran variedad de elementos de carácter afectivo, intencional y fenomenológico que es preciso tener en consideración (Kim, 2006: 14-17; Pérez, 2013: 44-59; Pineda, 2012: 46-54). En el derecho estos elementos están presentes de diversas maneras, a veces es necesario atribuir el conocimiento de ciertas circunstancias a una persona (*e.g.*, el conocimiento de ciertas normas, de su parentesco con alguien o de que un arma se encontraba cargada), a veces ciertas emociones (*e.g.*, aquellas que pueden excusarle de un castigo), e incluso sensaciones (*e.g.*, haber escuchado un disparo).

En segundo lugar, es común identificar diferencias entre los reportes y atribuciones de estados mentales realizados desde la primera persona y los realizados desde la perspectiva de la tercera persona[6]. Esto influye en la forma en que es posible atribuir estados mentales correctamente. Cómo las diferentes perspectivas utilizan elementos de diversa índole para atribuir estados mentales se puede ver con un ejemplo. Si Juan nos cuenta que fue a ver una película de terror y al preguntarle si sintió miedo nos dice que no, le creeremos. Nuestro juicio puede cambiar si alguien nos dice que cualquier persona que ha visto esa película siente miedo, así como si nos dijera que vio a Juan salir pálido del cine. Esto puede cambiar nuevamente si, confrontando a Juan, nos aclara que estaba pálido porque sentía un dolor estomacal. De todas formas, si sabemos que Juan es muy miedoso y que suele mentir acerca de estos asuntos, posiblemente optemos por no creerle.

En este contexto, en la tradición moderna suele entenderse a la introspección como un tipo de acceso a los propios estados mentales distinto al que terceros pueden tener, tales como la observación del comportamiento o el uso de la imaginación. Para muchos, dicha diferencia genera una especie de asimetría sobre el acceso a estados mentales, pues el acceso desde la primera persona sería directo, mientras que el acceso desde la perspectiva de la tercera persona sería indirecto. Esto se suele sostener bajo la idea

---

6 Es común dejar fuera de la discusión aquellos elementos en que entra en juego la segunda persona (véase Pérez, 2013, y, en el ámbito de la moral, véase Darwall, 2006). Dicha actitud ante este tipo de perspectiva me parece injustificada. De todas formas, no profundizaré en ello, aunque lo que se defenderá en las páginas finales de este texto muestra la relevancia de dicha perspectiva.

de que lo mental es algo *interno* de los individuos a lo que cada individuo tendría un acceso privilegiado, tesis muy difundida en la filosofía moderna.

Más allá de lo plausibles o descabelladas que puedan resultar las ideas presentes en el párrafo anterior, y considerando que en la práctica jurídica, por regla general, se es indiferente a los reportes de estados mentales cuya fuente es la introspección (*i.e.*, para realizar una correcta atribución de la comisión de un ilícito con dolo no es suficiente preguntarle al acusado si ha actuado dolosamente, así como para atribuir haber actuado con consentimiento al momento de contratar no basta con que la persona nos diga en qué estado mental se encontraba al momento de firmar), en lo que sigue se tendrá presente de forma general la posibilidad de atribuir estados mentales a otros. Más específicamente, lo que se dirá está guiado por las siguientes preguntas: ¿de qué forma otros individuos nos muestran o nos pueden mostrar lo que hay en su mente?, y ¿cómo es posible entender que tenemos acceso a ella? Así, esta reflexión tiene que ver más con la atribución de estados mentales que con la naturaleza de los mismos, por lo que puede entenderse como una reflexión acerca de qué supone dicha atribución, antes que sobre la naturaleza de lo mental. De todas formas, como se verá, ambas cuestiones no están completamente separadas, y lo que se afirme acerca de una dependerá parcialmente de lo que se afirme respecto de la otra, por lo que las preguntas referidas nos dirigen a otras tales como: ¿qué está supuesto en la idea de que podemos acceder a los estados mentales de otros?, y ¿cómo la conformación de lo mental nos permite decir con sentido que atribuimos estados mentales correctamente?

La visión estándar acerca de lo mental en las últimas décadas se puede entender en parte como un intento de articular dos ideas comúnmente aceptadas: la eficacia causal de lo mental y la primacía causal de lo físico (Corbí y Prades, 2000: cap. 1; Lowe, 2008: cap. 1). La primera se vincula al denominado problema mente-cuerpo, esto es, a la necesidad de explicar las interacciones que podemos identificar entre nuestros estados mentales y nuestro cuerpo y su entorno (*e.g.*, romperse un brazo conlleva la aparición de dolor y el deseo de comer conduce a caminar hacia la cocina en busca de alimento). Gran parte de nuestra vida mental, si no toda, tiene que ver con la manera en que enfrentamos al mundo con nuestro cuerpo. En este punto, las teorías contemporáneas de lo mental surgen como respuesta al dualismo de sustancias. Según está teoría, defendida de forma célebre

por Descartes en *Meditaciones metafísicas* (1642) y en el texto póstumo *Tratado del hombre* (1662), existirían dos tipos distintos de sustancias, *res cogitans* y *res extensa*, que se corresponderían con lo mental y lo físico respectivamente, vinculándose esta última al cuerpo en el caso de los animales[7]. En cuanto sustancias distintas, *res cogitans* y *res extensa* tienen diferentes características. Por ejemplo, solo lo físico tiene extensión, pero solo lo mental piensa. Más aún, al ser sustancias distintas existen de forma independiente la una de la otra, siendo concebibles diferenciadamente, por lo que son irreducibles la una a la otra.

Uno de los principales problemas para el dualismo de sustancias surge cuando se desea dar una explicación de cómo ambas sustancias interactúan, específicamente en la relación mente-cuerpo. Los movimientos corporales de las personas son uno de los principales medios con que se cuenta para acceder desde la tercera persona a los estados mentales de otros, como se puede ver de los ejemplos señalados más arriba. Para el dualismo de sustancias sería imposible dar cuenta de la idea de que la mente *causa* ciertos cambios en el mundo físico del cual el cuerpo forma parte (*i.e.*, la denominada causación mental). En la acción, noción central para comprender la agencia, los movimientos corporales característicos parecen estar precedidos por un estado mental que determina la forma en que el cuerpo interactúa con su medio (*e.g.*, desear comer algo determina los movimientos corporales en que consiste caminar hacia la cocina)[8].

Respecto a las relaciones causales, es una idea compartida que todo evento ocurre debido a que ha ocurrido otro u otros que son su causa, más allá de que seamos capaces o no de conocer completa y detalladamente todos los elementos de dicha relación entre eventos. Se trata de una relación fáctica que permite explicar y predecir eventos. Si aceptamos que lo mental

---

7 Haré referencia a animales sin distinguir entre animales humanos y no humanos. Considerando que en el ámbito jurídico los estados mentales relevantes son de animales humanos, en lo que sigue, salvo que se exprese lo contrario, lo que se predica sobre animales, se predica sobre seres humanos.

8 Lo señalado asume una visión causalista de la acción, visión dominante desde la aparición de los trabajos de Donald Davidson en la década de los sesenta (véase la compilación Davidson, 1980). Sobre esto se volverá más adelante.

existe como algo distinto de lo físico, como hace el dualismo de sustancias, entonces no es posible explicar su poder causal, pues solo se puede atribuir dicho poder a aquello que forma parte de lo físico[9]. Detrás de esta última premisa se puede identificar lo que Josep Corbí y Josep Lluís Prades denominan fisicalismo causal. Según esta doctrina, cualquier eficacia causal que se le pueda atribuir a una propiedad no física en un caso particular no puede ir más allá de los poderes causales de las propiedades físicas presentes en dicha ocasión particular. La doctrina es fisicalista porque asume que no hay nada fuera o sobre lo físico[10]. Traducido a nuestro ámbito de discusión, con esta doctrina se dice que, si se desea considerar a lo mental como *algo* constituido por propiedades no físicas, debe poder mostrarse que cualquier causación mental es dependiente de un proceso físico que le subyace, de lo contrario no será posible entenderla como causalmente relevante (Corbí y Prades, 2000: xiii, 2)[11]. Esto ha conducido a desarrollar una visión según la cual cada estado mental particular se corresponde con una instanciación de ciertas propiedades físicas en los individuos, cuestión central para la adopción de los pasos primero y segundo[12].

---

9 La otra consecuencia negativa que suele identificarse en el dualismo de sustancias se vincula con la tesis de Descartes según la cual tenemos un acceso directo a nuestros propios estados mentales; si bien algo se dirá más adelante, no es el objeto de este trabajo revisar esta problemática. Para una clásica revisión crítica de esta tesis véase Ryle, 1949: cap. 1; Kenny, 1989: cap. 1.

10 Sobre el fisicalismo y sus variables, véase Tiehen, 2018. Detrás de estas ideas se encuentra la tesis de la clausura causal del universo (véase Lowe, 2008: 41-78).

11 Un modelo que asume estas ideas puede verse en la propuesta de William Lowe (2008), según la cual eventos no físicos pueden formar parte de explicaciones causales.

12 La tesis de la identificación de propiedades mentales con propiedades físicas es común a diversas teorías sobre la naturaleza de lo mental. En el párrafo se hace referencia a instanciaciones (como traducción del inglés *token*) considerando los problemas de las denominadas teorías de la identidad de tipos (*type*), en particular su imposibilidad para explicar el así llamado principio de la múltiple realizabilidad. Esta dificultad se ha tratado de evitar con propuestas como el monismo anómalo, así como con diversos tipos de funcionalismo empírico y el conexionismo (véase Pineda: cap. 4-6).

2. NEUROCIENCIAS, MENTE Y CEREBRO

Esta conjunción de ideas, vinculadas a la necesidad de explicar la relación entre mente y cuerpo evitando el dualismo y, al mismo tiempo, a la defensa de que todo lo que tiene eficacia causal debe dicha eficacia a la instanciación de propiedades físicas, ha encontrado en las neurociencias un fuerte aliado. Una forma básica de responder al problema del dualismo de sustancias consiste en negar que existan dos sustancias y que es suficiente para comprender al mundo en general, y a lo mental en particular, comprometerse con la sola existencia de *res extensa*, generándose así un monismo de sustancias fisicalista, que se define a sí mismo como la alternativa posible al dualismo de sustancias.

Este monismo de sustancias fisicalista requiere rastrear aquellos procesos físicos que corresponden a estados mentales, y precisamente la generación de métodos que permiten ubicar a estos en el cerebro abre la puerta a la satisfacción de dicho requerimiento. Si bien la intuición de que los procesos mentales están íntimamente conectados con el cerebro es de larga data, ha sido con el actual desarrollo de las neurociencias que se ha expandido la idea de poder *ver* la mente de otras personas mediante el estudio de su cerebro (FREEMAN, 2011: 1-6)[13]. El reputado neurocientífico Francisco Rubia da cuenta de esta relación al referirse a su disciplina afirmando:

> La mayoría de los neurocientíficos han renunciado ya a la vieja idea de que cerebro y mente, o cuerpo y alma, son de naturaleza distinta [...] No tenemos ninguna prueba de que existan en el cerebro operaciones que no se deban a impulsos que provienen del propio sistema nervioso. [...] La reciente producción de experiencias místicas, espirituales –con estimulación electromagnética del lóbulo temporal– apunta, a mi entender, a un apoyo a la teoría monista de que la separación entre materia y espíritu no es más que un producto cerebral, pero no real fuera de nuestro cerebro (RUBIA, 2009: 100-101)[14].

---

13 Se resalta la expresión "ver" por referir a una forma de acceder a los objetos mediante nuestro aparato perceptual, lo cual forma parte importante de la forma en que se origina conocimiento (válido) en el modelo de pensamiento cientificista.

14 Según BENNETT y HACKER, 2005, no es sino hasta la tercera generación de neurocientíficos que se adopta este punto de vista.

Las dos primeras partes de la cita dan cuenta de la renuncia al dualismo de sustancias a partir del uso de evidencia científica. En la segunda parte, además, se asume que la identificación de lo mental se hace mediante la detección de operaciones del sistema nervioso, lo cual probaría la hipótesis del monismo de sustancias fisicalista. La última parte de la cita muestra cómo, por medio de los métodos de la neurociencia, se puede identificar, e incluso *producir*, lo mental mediante el estudio y la intervención de los procesos físicos que se dan en el cerebro. Esto no solo se predica respecto de estados mentales que se vinculan con experiencias místicas, sino también respecto de las creencias que adoptamos sobre ello, que no serían más que *un producto cerebral*. En esta línea, la neurociencia como estudio del cerebro permitiría acceder a los estados mentales de las personas y, con ello, solucionar la señalada problemática con la certeza proporcionada por el método científico[15].

Estas ideas muestran cómo se da el primero de los pasos apuntados en la introducción, esto es, considerar que lo primario al hablar de lo mental es referir eventos y estados de cosas que se definen por ocurrir en el cerebro, entendido este como un complejo conjunto de células. Según esto, todo estado mental de una persona ocurre en el cerebro y puede rastrearse en él. Maribel Narváez resalta que, en este contexto, los avances realizados por la neurociencia cognitiva, en especial por medio del desarrollo de técnicas de neuroimagen, prometen "ver 'la ontología' de nuestra mente de una manera muy potente, de ver eso que siempre permaneció oculto en el espacio privado delimitado de nuestro cráneo" (Narváez, 2016: 102-103)[16]. La cita refiere al rol que juega, dentro de los debates en torno a los conceptos mentales, la posibilidad de *ver el interior*, lo cual conlleva romper la supuesta asimetría entre los reportes de primera y tercera persona que algunos derivan de las diferencias que hay entre las diversas perspectivas desde las cuales se pueden atribuir estados mentales. Así, si Juan nos miente

---

15 Otro aliado dentro de las corrientes filosóficas es el naturalismo (véase Pardo y Patterson, 2013: 25; Reiner, 2011: 164-165). Un interesante grupo de artículos acerca de las posibilidades del naturalismo en el conocimiento jurídico en Brożek y Stelmach, 2017.

16 La autora citada considera esto como una "ilusión".

al decirnos que no sintió miedo al ver la película, tenemos la posibilidad de descubrir si nos dice la verdad o nos miente, pues bastaría con contar con una IRMf (imagen por resonancia magnética funcional) de su cerebro en el momento en que veía la película. Estos hallazgos podrían incluso devenir en una inversión de la asimetría, pues muchas veces los individuos se autoengañan, no son conscientes de sus estados mentales o confunden unos estados mentales con otros, problemas que no tendrían los métodos de neuroimagen, los cuales presentarían de forma transparente, a terceros, lo que ocurre en la mente (*i.e.*, en el cerebro). Por supuesto, esto se debe entender como buenas noticias, en especial para el ámbito jurídico, pues al convertirse la mente de un individuo en algo transparente para terceros se podría, por ejemplo, acceder directamente a la aceptación de una norma por parte de un individuo, a su intención dolosa en la comisión de un ilícito o a su consentimiento al celebrar un contrato. De este modo, el neuroderecho aparece como una disciplina muy prometedora[17].

Estas consideraciones nos llevan al tercer paso, pues, de la hipótesis de que lo que usualmente llamamos una creencia, la aceptación de una norma o tener miedo no son más que un conjunto de relaciones que se dan entre células en determinada parte del cerebro, se concluye que la agencia de las personas se explica a partir de ese conjunto de eventos o estados de cosas definidos por sus propiedades fisiológicas, toda vez que la agencia se constituye a partir de la tenencia de ciertos estados mentales que definen lo que es deliberar, tomar decisiones y actuar conforme a ello. Lo que entendemos comúnmente como una creencia y que nos lleva a atribuirla en ciertas condiciones y en otras no, es más bien un mito (o el producto de una mala teoría), y lo que *realmente* explica lo que ocurre son relaciones físicas entre células[18].

---

17 En este punto estoy obviando posibles conflictos valorativos que pueden derivar en la afectación de derechos fundamentales por medio de la utilización de estas técnicas.

18 Usualmente es en este punto en donde algunos autores dan un cuarto paso y afirman cuestiones como: "el libre albedrío es una ilusión cerebral" (Rubia, 2009: 109). Para algunos neurocientíficos, de esto se sigue que lo que realmente define nuestro comportamiento son ciertos procesos físicos de los cuales no tenemos plena conciencia y que, en el mejor de los casos, tratamos posteriormente de explicar por medio de elucubraciones (véase Libet, 1985; Wegner, 2003). Es razonable pensar que en este

Considerando la importancia para el derecho de lo que implica atribuir estados mentales, vale la pena detenerse en la idea de que en este marco lo mental se predica del cerebro. Es precisamente en esto en lo que consiste el segundo de los pasos acusados en la introducción[19]. Se trata de una idea bastante extendida y que se puede encontrar en diversas fuentes. Así, por ejemplo Francis Crick, en *The Astonishing Hypothesis*, señala:

> Lo que ves no es realmente lo que hay; es lo que tu cerebro cree que hay [...] El cerebro hace la mejor interpretación que puede de acuerdo con su experiencia previa y la información limitada y ambigua que los ojos le proporcionan [...] El cerebro combina la información que las muy diferentes características de la escena visual ofrecen (aspectos de la forma, el color, el movimiento, etc.) y se asienta en la interpretación más verosímil de todos esos indicios tomados en su conjunto [...] Lo que el cerebro debe construir es una interpretación de múltiples niveles de la escena visual [...] [Este proceso de reconstrucción] permite al cerebro conjeturar una imagen completa a partir de solo información parcial – una habilidad muy útil (citado por Bennett y Hacker, 2005: 68; 2007: 29-30)[20].

Para M. R. Bennet y P. M. S. Hacker, asumir esto nos lleva a entender que "el cerebro tiene experiencias, cree cosas, interpreta pistas a partir de la información que se le facilita y hace conjeturas" (Bennett y Hacker, 2007: 30; 2005: 68-70). Lo que se defiende con esto es la tesis de que la mente *está en* o *es* el cerebro, llegando a afirmarse la existencia de una

---

modelo dichas elucubraciones, en cuanto estados mentales, también deben tener un correlato fisiológico. Este camino llevaría a un regreso al infinito. Este problema se ha intentado solucionar presentando teorías de la superviniencia, en donde los estados mentales supervienen a procesos y estados físicos, así como por medio de la comprensión de los estados mentales como categorías funcionales (véase, *inter alia*, Tiehen, 2018).

19 La relación de esta tesis con las neurociencias ha dado paso al denominado *neuroesencialismo*, el cual se entiende como "la visión según la cual la forma definitiva de explicar la experiencia psicológica humana es haciendo referencia al cerebro y su actividad desde perspectivas químicas, biológicas y neurocientíficas" (Schultz, 2015: 2. Véase Reiner, 2011).

20 Afirmaciones similares en Rubia, 2015: 15-16, 27.

identidad entre ambos que nos habilitaría para realizar ese tipo de juicios acerca del cerebro. En palabras de Narváez:

> Las investigaciones sobre el desarrollo, la estructura y el funcionamiento del cerebro parecen establecer correlaciones entre cómo *está/es* el cerebro en un determinado momento y *qué es lo que hacemos*. No se trata del cerebro en su conjunto sino, según los casos, de una concreta zona de éste, un tipo de conexiones, el estado químico en un área específica, o caudal del flujo sanguíneo, por ejemplo.

Pero la cuestión no termina en este punto, pues, como señala la citada autora:

> Una correlación, por supuesto, no distingue entre *se hace x porque así está el cerebro* o *así está el cerebro porque se hace x*. Huelga decir que esas no serían las dos únicas formas de interpretar una correlación. Cuando prima el valor causal del estado cerebral en tales correlaciones, la capacidad explicativa de la correlación se extiende en el tiempo, tanto hacia atrás –hizo x porque el cerebro estaba así– como hacia adelante –hará x cuando el cerebro se encuentre así– (Narváez, 2016: 107).

Esta es la forma en que se incorpora la relación causal en la explicación de lo mental, estableciéndose una relación cerebro-mente en que no solo hay correlación, sino una necesidad fáctica. La relación se convierte en una en que la identificación de eventos y estados de cosas en el cerebro, cuyas propiedades características son fisiológicas, es condición necesaria y suficiente para identificar y explicar un estado mental, lo que para algunos supone defender la tesis de que *es* el cerebro el que piensa, cree, teme, etc. Una correcta atribución de estados mentales consiste en la atribución de estados al cerebro.

3. DUALISMOS Y LA VIDA DE LOS SIGNOS

Considerando lo señalado en los párrafos anteriores, es posible reconocer que parte del pensamiento contemporáneo acerca de la mente identifica

lo mental con lo cerebral[21], lo que lleva a, por ejemplo, predicar estados mentales del cerebro, así como a pensar la agencia en términos de lo que sucede en nuestro sistema nervioso. Como se señaló al comienzo de este trabajo, vale la pena mirar con recelo los tres pasos y las conexiones entre ellos que conducen a ver las cosas así. Una primera razón para ello, sobre la que no me extenderé, es que las teorías que se sustentan en esto no logran hacer la conexión que se supone que era posible explicar al rechazar el dualismo, pues al generar una identidad entre mente y cerebro aún queda por explicar cómo el cerebro interactúa con el cuerpo (véase, *inter alia*, Corbí y Prades, 2000; Crane, 1995).

Una segunda cuestión, más importante para este trabajo, es que el modelo reduccionista genera otro tipo de dualismo, manteniendo una parte relevante del argumento cartesiano, en especial cuando se habla de agencia. Esto lleva a que Lynne Rudder Baker denomine a la visión estándar como *cartesianismo materialista* y Bennett y Hacker como *neocartesianismo*. El punto central es que, si bien desde el monismo de sustancias fisicalista se rechaza el dualismo de sustancias, se sigue pensando la relación mente-cuerpo de forma dualista, toda vez que se asume que predicar cosas de la mente supone predicar cosas de algo que es distinto al individuo que actúa, haciendo una escisión entre el agente y la (su) mente. En este caso se predican cosas del cerebro, siendo los comportamientos del agente un mero efecto de lo que ocurre en aquel.

En consecuencia, un supuesto compartido por la tercera generación de neurocientíficos y los filósofos en que esa generación se apoya, por una parte, y el dualismo de sustancias, por otra, es que los atributos psicológicos son algo que se predica de lo mental y no de los animales en cuanto individuos. En el caso de Descartes estos atributos se predican de la *res cogitans*, en el caso de los monistas de sustancia se predican del cerebro.

Otro supuesto compartido es considerar a la mente como una cosa que interactúa con el cuerpo, lo que lleva a separar la mente del comportamiento

---

21 Según Pardo y Patterson, 2013: 7, este reduccionismo toma dos formas: se acepta una identidad según la cual el cerebro *es* la mente o se acepta una prioridad explicativa según la cual los atributos mentales pueden ser explicados por completo en términos de información acerca del cerebro.

y, con ello, de las acciones. En el caso de Descartes la explicación de esta interacción le llevó a desarrollar teorías muy criticadas acerca del funcionamiento de la glándula pineal; en el caso de los monistas de sustancia esta interacción se da entre dos cosas que forman parte de la *res extensa*, por lo que la cuestión no parece ser tan problemática. Para estos últimos, ciertos estados en que se encuentra el cerebro causan los movimientos corporales, y el acceso a lo mental se hace por medio de técnicas de neuroimagen y solo indirectamente a través del comportamiento de los individuos. En este marco, el comportamiento que forma parte de las acciones pasa a ser un signo de la vida mental que se da en el cerebro, un efecto de lo que ocurre en este[22]. El argumento tiene la misma estructura que el de Descartes pues, a pesar de que se niega la existencia de distintas sustancias, se hace la distinción entre el animal que realiza ciertos comportamientos con su cuerpo y el cerebro de dicho animal (que vendría a reemplazar a la *res cogitans* en el argumento). Lo mental se predicaría del segundo y solo derivativamente del primero, manteniéndose la dualidad (véase Bennett y Hacker, 2005: 62-65, 72-73; Hacker, 1997: 16; Narváez, 2016: 110-114; Pardo y Patterson, 2013: 44; Rudder Baker, 1995: 6-7).

Si consideramos la interpretación que hacen Barry Stroud y David Finkelstein de las observaciones de Ludwig Wittgenstein respecto al seguimiento de reglas (véase Finkelstein; 2008: 38-44, 79-88; Stroud, 1996), podemos identificar un problema filosófico de fondo que acompaña a este dualismo acerca de lo mental (dualismo compartido por monistas de sustancias y dualistas de sustancias). Según Finkelstein, para identificar parte de los problemas asociados con el seguimiento de reglas en el ámbito de lo mental, podemos acudir a la relación que se establece entre una orden y su ejecución. Así, en la sección 431 de *Investigaciones filosóficas* (2009) Wittgenstein pone en la boca de su interlocutor las siguientes palabras:

---

22 Esto nos lleva a una tercera cuestión que tienen en común con el modelo de Descartes: la distinción entre lo interior y lo exterior, donde lo mental se encuentra *dentro* del individuo, ya sea en su alma, ya sea en su cerebro (véase Rudder Baker, 1995: 14; Schatzki, 1996: 40-41).

> "Entre una orden y la ejecución hay un abismo. Este tiene que ser cerrado mediante la comprensión".
>
> "Solo ante la comprensión se dice que tenemos que hacer ESO. La *orden* –eso son meros sonidos, trazos de tinta".

Según Stroud y Finkelstein, un error común dentro de la filosofía moderna consiste en aceptar, junto al interlocutor de Wittgenstein, que existe un abismo entre la orden y su ejecución a partir de la dificultad que puede surgir al momento de identificar la forma en que debe cumplirse una orden.

La misma dificultad surgiría al intentar descubrir o atribuir significado a cualquier texto escrito o a los sonidos que emanan de la boca de una persona, habría un abismo entre dichos sonidos o marcas de tinta y el significado. Una vez que aceptamos dicho abismo, asumimos que una secuencia de palabras escritas o dichas no puede ser entendida sin algún tipo de acto que sirva de *puente* entre la tinta o el sonido y el significado, o entre la orden y su ejecución[23]. De esta forma, si es que hay un abismo entre cada orden y su ejecución, entonces siempre una orden puede ser mal interpretada y todo intento de comprensión funciona como una interpretación que "pende, juntamente con lo interpretado, en el aire; [y] no puede servirle de apoyo" (WITTGENSTEIN, 2009: § 198).

Según estos intérpretes de Wittgenstein, el punto de partida de este problema está en asumir que las palabras son simplemente manchas de tinta que deben ser interpretadas para asignarles significado o para *conectarlas*

---

23 En *Cuaderno azul*, Wittgenstein muestra esto señalando: "Se puede preguntar ahora: ¿interpretamos las palabras antes de obedecer la orden? Y se encontrará que en algunos casos se hace algo que podría llamarse interpretar antes de obedecer, y en otros casos no. Parece que hay ciertos procesos mentales definidos, vinculados con la actuación del lenguaje, procesos únicamente a través de los cuales puede funcionar el lenguaje. Quiero decir los procesos de comprensión y significación. Los signos de nuestro lenguaje parecen muertos sin estos procesos mentales; y podría parecer que la única función de los signos es inducir tales procesos, y que éstas son las cosas en las que deberíamos estar realmente interesados. Así, si se les pregunta cuál es la relación entre un nombre y la cosa que nombra, tenderán a contestar que la relación es una relación psicológica, y quizá al decir esto estén pensando particularmente en el mecanismo de la asociación" (WITTGENSTEIN, 1958: 30-31).

con su significado y, así, hacer un puente. En términos generales, esto se traduce en la idea de que siempre hay un abismo entre la formulación de una regla y su aplicación, pues cualquier sonido y cualquier mancha de tinta se pueden calzar con cualquier significado y con la realización de cualquier acción. Esto conduce a la más famosa formulación del problema que deriva del seguimiento de reglas, presente en el primer párrafo de la sección 201 de *Investigaciones filosóficas*, que expresa:

> Nuestra paradoja es esta: una regla no podría determinar ningún curso de acción ya que todo curso de acción ha de ponerse de acuerdo con la regla. La respuesta fue: Si todo se puede poner de acuerdo con la regla, entonces también con la contradicción. Por lo que no habría aquí ni concordancia ni contradicción.

Para Finkelstein algo similar sucede con la concepción de lo mental dominante en filosofía y que es común a ambos tipos de dualismo. Más específicamente, se entiende que hay un abismo entre el comportamiento humano y su significación psicológica que lleva a considerar que hay una sustancia distinta a la que corresponde lo mental (como ocurre con el dualismo de sustancias) o bien a reducir lo mental a una parte del cuerpo (como sucede con el monismo de sustancias). Ambas perspectivas comparten la idea dualista de que el comportamiento humano es simplemente un conjunto de movimientos corporales que se entienden como tales debido a su interacción con una mente, la cual es, a su turno, entendida como una *cosa* distinta.

El problema detrás de esto es acusado por el mismo Wittgenstein, quien señala en el segundo párrafo de la sección 201 de *Investigaciones filosóficas*:

> Que ahí hay un mal entendido se muestra en tanto que en este razonamiento suponemos interpretación tras interpretación; como si cada una nos tranquilizase al menos por un momento, hasta que pensamos en una interpretación que de nuevo está detrás de ella. Con ello mostramos que hay una concepción de una regla que *no* es una *interpretación*, sino que se manifiesta, de caso en caso, en la aplicación, en lo que llamamos "seguir la regla" y lo que "la contraviene".

El problema se encuentra precisamente en asumir que para entender una regla es necesario siempre acudir a una interpretación de la misma o a algún otro recurso que cumpla la función de puente. Esto es así porque de ello surge, tarde o temprano, la necesidad de conectar dicha interpretación con su significado, identificándose un abismo entre la interpretación y lo interpretado que requiere, a su vez, de una interpretación, y así sucesivamente, generándose un regreso al infinito[24]. Pero el punto para Wittgenstein es precisamente que cuando el problema no está presente es porque no hay una interpretación y somos capaces de identificar sin dificultad las acciones que siguen o contravienen el estándar que se aplica. Lo mismo puede decirse del seguimiento de órdenes o de la comprensión de un texto.

La alternativa que se inspira en las ideas de Wittgenstein consiste en rechazar la idea de que las palabras no son más que sonidos o manchas de tinta sobre un papel y que hay un abismo entre ellas y su significado. En este punto vale la pena acudir a la sección 432 de *Investigaciones filosóficas* en donde se lee:

> Todo signo parece *por sí solo* muerto. ¿*Qué* le da vida? – En el uso *vive*. ¿Tiene ahí la respiración vital en sí? – ¿O es el uso su respiración?

El signo (manchas de tinta, sonidos, movimientos corporales) solo aparece como muerto cuando lo vemos fuera de su uso, cuando lo vemos *por sí solo*, fuera de las actividades humanas que le dan vida. Pero, como Wittgenstein propone en el segundo párrafo de la sección 201, la problemática se esfuma cuando notamos que la regla se manifiesta en su aplicación a conductas que la contravienen y que la siguen; en ese momento la idea de que se está realizando una interpretación pierde sentido y la regla cumple su papel en cuanto regla[25].

---

24 Los esfuerzos por crear un puente entre la formulación de una regla y su aplicación usualmente han fluctuado entre lecturas realistas y antirrealistas, por un lado, y propuestas platónicas y convencionalistas, por otro (véase DUMMETT, 1959; STROUD, 1965). Una selección de algunos de los principales textos que tratan el tema del seguimiento de reglas en MILLER y WRIGHT, 2002.

25 Es un lugar común en la teoría de las normas decir que guían la conducta, tal como una orden lo puede hacer (al menos, cuando es exitosa). Pensar que una orden o la

El uso de un signo depende del contexto en que funciona como tal. A pesar de que a veces nos es difícil entender el significado de un enunciado o de una frase (y, por ende, requerimos interpretarlas) y que a veces estamos en desacuerdo con otros acerca de ello, podemos hacer la distinción entre manchas de tinta en un papel y palabras con significado precisamente porque una mancha de tinta tiene sentido como algo que debe ser interpretado solo una vez que conocemos y estamos familiarizados con el alfabeto escrito (o con los sonidos que otros emiten, o con su conducta) y somos capaces normalmente de entender lo que con él se expresa sin necesidad de realizar interpretaciones[26]. Podemos identificar algo como una palabra o unas manchas de tintas como letras solo porque ellas tienen vida dentro de actividades humanas. Tal como normalmente, cuando leemos un libro, no necesitamos interpretar palabra por palabra lo que está escrito, cuando se nos da una orden simplemente la ejecutamos, no buscamos una interpretación para saber qué hacer, y algo es una orden precisamente porque debe ser ejecutada. Esto no niega que de vez en cuando sea necesario realizar dicha interpretación, pero sí significa que es descabellado asumir que todo acto de comprensión es una interpretación.

Normalmente nada hace de puente en el abismo entre unas manchas y un significado debido a que normalmente no hay tal abismo. En palabras de Filkenstein:

> Wittgenstein constantemente nos recuerda que el fenómeno en el que él está interesado –proveer interpretaciones, definir ostensivamente, realizar un movimiento en un juego de ajedrez, esperar una explosión, sentir dolor– tiene sentido 'únicamente en medio de ciertas manifestaciones normales de la vida' (*Zettell*, 534).

---

formulación lingüística de una regla siempre requiere de una interpretación pierde de vista esto y convierte una cuestión práctica en una cuestión intelectual o reduce las reglas a las intenciones del intérprete. Sobre este punto, véase Narváez, 2015.

26 Cabe señalar que para esta tesis es indiferente si se defiende un innatismo sobre conceptos o que la adquisición de conceptos se realiza por medio de educación, pues no se trata de dar cuenta del origen de la adquisición de dichos conceptos. Lo mismo cuenta para la afirmación, que se hará más adelante, sobre la posibilidad de identificar *expresiones naturales* de estados mentales.

En cuanto intentemos "desvestir" palabras –despojarlas de sus contextos y entenderlas como garabatos– seremos incapaces de dar sentido a la sugerencia de que "hay una concepción de la regla que no es una interpretación" (FILKENSTEIN, 2008: 82). Desvestir a las palabras o la conducta viéndolas como meros signos se traduce en una comprensión de lo mental según la cual el comportamiento de otros aparece como algo inerte que hay que conectar con aquello que le da vida. Esto supone ver a la conducta como una serie de movimientos corporales y a los reportes de estados mentales como sonidos que deben ser interpretados de tal manera que refieran a algo que los hace verdaderos en cuanto partes de la vida mental, se encuentre aquello a lo que se hace referencia en la *res extensa* o en la *res cogitans*. Precisamente esto conduce a asumir un dualismo, donde la conducta humana es solo un signo cuyo significado, en cuanto estado mental, se encuentra en otro objeto, sea el alma o el cerebro. Un riesgo de ello es que la conducta de las personas se vuelve irrelevante, pues bastaría con las técnicas de neuroimagen para conocer lo que es relevante conocer. Además, el comportamiento de un individuo aparece como un mero efecto en el mundo de algo que ha sucedido en el cerebro (o en el alma).

Wittgenstein no estaría presentando una teoría acerca de cómo hacer un puente para terminar con el abismo, ni una teoría escéptica acerca de que haya algo que pueda hacerlo[27], lo que estaría mostrando es que una mala concepción de lo que es una regla (una palabra, la conducta humana) nos lleva a asumir dicho abismo. A su vez, nos muestra que procesos como los de interpretación tienen sentido sobre un trasfondo de casos en los cuales no se plantea duda sobre lo que dice una regla (o una orden, o un movimiento corporal). De esta forma, puede surgir una tercera opción acerca de cómo comprender lo mental; en la siguiente sección se explorará en qué puede consistir.

---

27 Esta ha sido una de las lecturas dominantes sobre la cuestión, especialmente desde la publicación del ensayo de Saul Kripke *Wittgenstein on rules and private language* en 1982.

4. EXPRESIÓN, MENTE Y CONDUCTA

Michael Pardo y Dennis Patterson señalan que la visión estándar reduccionista presentada en la sección II se ve a sí misma dentro de una dicotomía entre un dualismo cartesiano y un reduccionismo de lo mental a lo cerebral que se sigue a partir de un monismo de sustancias fisicalista[28]. Para Pardo y Patterson dicha dicotomía es algo aparente, pues es concebible otra forma de entender lo mental que no responde a dicha distinción. Para acceder a esta opción es necesario tomar como punto de partida un grupo distinto de ideas de las antes vistas.

Dos tesis son particularmente relevantes. La primera de ellas es negar uno de los supuestos del dualismo compartido entre cartesianos y reduccionistas, afirmando que lo mental no es una cosa separada del individuo y su conducta. La segunda consiste en afirmar que lo mental es parte de la comprensión que tenemos de nosotros en cuanto animales y agentes, sin lo cual dicha comprensión carece de sentido. Según este segundo punto, lo mental se predica de los individuos en cuanto personas y no en cuanto entidades materiales, y procurar *ubicar* lo mental en alguna parte del cuerpo es un sinsentido. Tal como la habilidad de ver no es una parte del ojo que interactúa con otros elementos del ojo físico, ni la habilidad de volar puede ser entendida como una parte de un ave, las capacidades vinculadas a lo mental no deben aislarse del comportamiento humano y de la noción de la persona que se expresa a través de aquel.

En esta línea, autores como E. J. Lowe, Bennett y Hacker señalan que cuando se afirma que el cerebro piensa, decide o siente, simplemente se están exponiendo sinsentidos. No tenemos cómo saber qué quiere decir que un cerebro decida ir a la cocina a comer o que siente tristeza, pues son las personas las que deciden ir a la cocina y sienten tristeza. El problema de

---

28 Si se considera en el escenario la actitud eliminativista, desde una perspectiva cartesiana solo habría tres formas de entender los estados mentales: (a) o bien son estados cerebrales que tienen una existencia material espacio-temporal, (b) o bien son estados inmateriales, (c) o bien no existen los estados mentales. El reduccionismo es representado por (a), el dualismo de sustancias por (b) y el eliminativismo por (c). Véase RUDDER BAKER, 1995: 6.

este tipo de afirmaciones es que predican de una parte del animal lo que solo tiene sentido predicar respecto del animal como un todo (véase VON WRIGHT, 1997: 98)[29]. Por otro lado, debe considerarse que estos predicados no hacen referencia a una cosa, sino más bien a estados en los que los agentes se encuentran o a las capacidades y habilidades con que cuentan. Cuando se habla de los estados mentales como cosas (en este caso, procesos que ocurren en el cerebro) que se pueden describir, se utiliza una forma de hablar, basada en la relación objeto-designación, que deja de lado el sentido que tienen los conceptos mentales[30].

Una consecuencia de negar que la mente es una cosa en el sentido referido, es que surge el problema de saber cómo identificar estados mentales y, junto con ello, atribuirlos a otros. Se trata de un problema, porque cuando identificamos estados mentales con estados cerebrales basta con tener las herramientas para saber cómo se encuentra el cerebro de una persona para hacer una correcta atribución de un estado mental que se individualiza precisamente a través de dichos estados cerebrales. Cuando dicha tesis es cuestionada, se requiere de otros criterios positivos de identificación.

Una alternativa para resolver este problema supone incorporar la idea de expresión, entendida como una relación interna entre lo mental y el comportamiento. En el día a día, sabemos de los estados mentales de otras personas a través de su comportamiento (*e.g.*, salir caminando pálido del cine) y de sus reportes (*e.g.*, la afirmación del espectador de que la película no le asustó), los que se dan en contextos determinados en los cuales adquieren su sentido como tales. Podría decirse que ese sigue siendo el material básico con el que trabajamos. Por el contrario, como se señaló en la sección anterior, cuando consideramos al comportamiento humano solo como movimiento corporal que necesita ser interpretado, caemos en

---

29 Para Bennett y Hacker, decir que lo mental se predica de parte de la persona (*i.e.*, su cerebro o áreas del mismo) y no de la persona supone caer en la falacia *mereológica* (véase BENNETT y HACKER, 2005: cap. 3; PARDO y PATTERSON, 2013: 20-22, 44, 76-77, 99-105). Una doctrina como la de Bennett y Hacker encuentra antecedentes en la propuesta de Aristóteles (véase *De anima*: libro I).

30 Sobre las observaciones de Wittgenstein respecto de este punto, véase WILLIAMS, 1999: cap. 1.

la necesidad de incorporar un acto (o hecho, o experiencia) que medie entre dicho movimiento corporal y lo mental y, de paso, lo despojamos de su contexto.

La noción de expresión puede contrastarse con la de descripción. La idea de descripción de lo mental se utiliza cuando, asumiendo el fisicalismo monista, se dice que un estado mental se puede *percibir* por medio de los mecanismos de neuroimagen o, por introspección, por medio del *ojo* de la mente[31]. En esta línea, se entiende que una creencia es un cambio en la irrigación sanguínea de cierta zona del cerebro en conjunto con la activación de un grupo de neuronas, por lo que puede ser perfectamente descrita como tal. En contraste, el *acceso* a lo mental para quien toma en cuenta la noción de expresión proviene de la interacción con otras personas, al enfrentar su comportamiento significativo en cuanto otras. Así, cuando un niño se lastima y grita no está informando a su madre acerca de cierto estado de cosas al que accede por medio de introspección, ni la madre ve lo que ocurre en el niño como el efecto de la activación de ciertas zonas del cerebro (véase HACKER, 1997: 31-39; SCHATZKI, 1996: 25-30). De hecho, una respuesta como "qué interesante" o una como "¿Cómo sabes que te duele?, tal vez si te tomamos una IRMf podríamos ver si realmente estás sintiendo dolor" por parte de la madre están fuera de lugar; por el contrario, normalmente el niño manifiesta dolor y la madre acude a consolarle. Es en este tipo de reacciones en donde estaría el origen de nuestros juegos de lenguaje, no en descripciones de estados *internos* o de reacciones químicas[32]. Si, por una parte, cuando decimos "me duele la muela" o "estoy pensando" no estamos haciendo referencia a una parte de nuestro cuerpo, sino a nosotros, y si, por otra, en condiciones normales no hay un abismo

---

31 Esta idea de que accedemos a nuestros estados mentales por medio de un sentido especial no solo pertenece a la tradición cartesiana, sino también se encuentra en parte de la neurociencia contemporánea cuando se propone entender a la conciencia como un truco que nos juega la mente y que surgiría luego de percibir las verdaderas causas de nuestro comportamiento (véase WEGNER, 2003).

32 PARDO y PATTERSON, 2013: 32, consideran que esto también puede dar respuesta a la idea de que la psicología popular es solo una teoría con su propio vocabulario, teoría defendida por eliminativistas, aunque acogida de diversas formas por autores como David Lewis y Jerry Fodor (véase RUDDER BAKER, 1995: cap. 3).

entre esa oración y su significado, podemos afirmar que la relación entre reporte y estado mental no es la de una descripción, sino una de expresión.

Que se trate de expresiones manifiesta que hay una relación interna entre mente y conducta. La evidencia primaria para atribuir predicados psicológicos a un individuo es precisamente su comportamiento dentro de un contexto. Bennett y Hacker argumentan que no se trata de evidencia inductiva, entendida esta como la correlación de fenómenos concomitantes. Los criterios para atribuir este tipo de predicados están conceptualmente conectados con el atributo psicológico en cuestión. El dolor y el comportamiento de dolor no están correlacionados como dos hechos que coinciden; que las personas se quejen cuando sienten dolor no es una relación que se descubra empíricamente, como sí lo es descubrir que cierto estado neurofisiológico se repite cuando una persona siente dolor (véase Bennett y Hacker, 2005: 82-85)[33].

Al respecto, en la sección 244 de *Investigaciones filosóficas* se puede leer:

> ¿Cómo se *refieren* las palabras a las sensaciones? – En eso no parece haber ningún problema; ¿pues no hablamos diariamente de sensaciones y las nombramos? ¿Pero cómo se establece la conexión del nombre con lo nombrado? La pregunta es la misma que esta: ¿cómo aprende un hombre el significado de los nombres de sensaciones? Por ejemplo, de la palabra 'dolor'. Esta es una posibilidad: las palabras se conectan con la expresión primitiva, natural, de la sensación y se ponen en su lugar. Un niño se ha lesionado y grita; y ahora le tranquilizan los adultos y le enseñan exclamaciones y más tarde oraciones. Ellos le enseñan al niño una nueva conducta de dolor.
>
> "Dices, pues, que la palabra 'dolor' significa realmente el gritar?". – Al contrario; la expresión verbal del dolor reemplaza el grito y no lo describe.

Considerando lo dicho, Wittgenstein presenta como alternativa identificar *expresiones primitivas* de estados mentales. En un sentido elemental, este

---

33 Cabe aclarar que los funcionalistas analíticos, según la terminología utilizada por David Pineda, están dispuestos a aceptar esta relación entre estados mentales y procesos neurológicos identificada por inducción (véase Pineda, 2012: cap. 5; Rudder Baker, 1995: 7-20).

tipo de manifestaciones permiten comprender que el comportamiento intencional es expresivo, cuestión que se puede ver también en animales no humanos. Tanto en la interacción entre ellos como en su interacción con nosotros, su comportamiento es expresivo de su mente, formándose lo que Finkelstein denomina el espacio lógico de la vida animal (véase FINKELSTEIN, 2003: cap. 6; von WRIGHT, 1997: 100-101). A su vez, las reacciones al comportamiento son parte del mismo espacio (véase HACKER, 1997: 36-38, 49; SCHATZKI, 1996: 61-64; STROUD, 1996: 178-180). El comportamiento de un cazador tras su presa es expresivo de su intención de cazarla y la presa así lo percibe, por lo mismo puede entenderse su conducta como (expresiva de) su huida. Expresar lo mental por medio del propio comportamiento y comprender el comportamiento de otros como expresión de lo mental es parte de pertenecer al mundo animal, al mundo de lo mental; no hay nada *tras* eso que nos *haga creer* que es así[34]. Llorar al sentir un dolor, reír al sentir cosquillas, estirar los brazos hacia un objeto al desearlo, son formas primitivas de expresión que parecen preceder nuestros juegos de lenguaje y que, a su vez, son roca sólida sobre la que se desenvuelven (véase ANSCOMBE, 1963: 68-69; HACKER, 1997: 36-39; SCHATZKI, 1995: 61-63).

Los conceptos mentales se adquieren por medio de la interacción con otros, en diversos procesos de aprendizaje. Una expresión tiene sentido como tal en gran medida porque puede ser entendida por otro. La incapacidad de un individuo de identificar un grito de dolor desgarrador como una expresión de dolor o una risa como expresión de alegría (o de nerviosismo, dependiendo del contexto), y así sucesiva y sistemáticamente, significa simplemente que ese individuo es incapaz de comprender a otros. Josep Lluís Prades argumenta que la constatación de esto permite entender qué está en la base de nuestras interacciones, pues un animal que sistemáticamente interpreta el comportamiento de otros de una forma alternativa no es

34 Recordemos que, según muchos pensadores, nuestro cerebro produce nuestros estados mentales conscientes, haciéndonos creer cosas. Un problema central para este tipo de lecturas que se sigue de lo acá dicho consiste en que solo puede realizarse una correlación entre un estado mental y un estado cerebral una vez que ya sabemos identificar por otros medios ese estado mental. No me extenderé en este punto.

siquiera imaginable, ya que su comportamiento no puede ser comprendido como la identificación de intenciones (véase PRADES, 2011: 72, 75).

Para situar al argumento desarrollado en los párrafos anteriores, cabe recordar que al referir a lo mental hablamos de un conjunto de cuestiones muy disímiles, tales como sensaciones, emociones y actitudes proposicionales, que a su vez se pueden entender como complejos conjuntos de cuestiones fenoménicas e intencionales, con las que usualmente se hace referencia a una serie de capacidades, poderes y habilidades que se atribuyen a los individuos. Esto hace necesaria la identificación y clasificación de diversos criterios para diferentes estados mentales teniendo en cuenta cuál de dichas cuestiones prima o si estos estados mentales tienen o no duración temporal, si pueden ser compartidos por varias personas al mismo tiempo, entre otros[35].

De hecho, no todo estado mental funciona de la manera descrita, pero esto no abre una brecha entre mente y comportamiento. Consideremos el estado mental *intención.* Si una persona nos dice "intentaré salir de este cuarto", esperamos por definición que en algún momento su comportamiento exprese dicha intención de salir. Algo similar se puede decir de las creencias: si una persona dice que cree que la sala está llena de gente para que sus palabras tengan sentido (y, por ejemplo, no se le acuse de actuar irracionalmente o de mentir) no puede luego actuar como si el lugar estuviera vacío. En esta línea, respecto de las creencias, en la sección 578 de *Investigaciones filosóficas* se lee:

> Pregúntate: ¿Qué significa *creer* en el teorema de Goldbach? ¿En qué consiste esta creencia?
>
> [...]
>
> Quisiera preguntar: ¿Cómo interviene esta creencia en este teorema? Examinemos qué consecuencias tiene esta creencia, a dónde nos lleva, "Me lleva a buscar una prueba de este teorema" – Bien; ¡Ahora examinemos en

35 Este punto se encuentra desarrollado en las observaciones de Wittgenstein sobre lo mental a las que dedicó sus últimas décadas de vida. Para una revisión detallada véase BUDD, 1993.

> qué consiste realmente tu búsqueda! Entonces sabremos en qué consiste la creencia en ese teorema.

Cuando un individuo cree algo está dispuesto a realizar determinados comportamientos y a afirmar determinadas cosas acerca de lo que hay y no hay; de este modo, atribuir a otro una creencia supone atribuirle los compromisos que se asumen en dicho contexto al afirmarse la posibilidad de lo creído. Entender que alguien tiene una creencia supone entender *a dónde lo lleva*. Por eso, cuando alguien nos dice que cree en algo, debemos preguntarnos qué consecuencias se siguen de dicho reporte[36]. A su vez, si vemos que una persona dice creer en algo, pero actúa de forma incompatible con dicha creencia, estamos autorizados a cuestionar la veracidad de su reporte, o a preguntarle si realmente sabe lo que significa creer en lo que dice que cree[37]. Prades señala que en la base de estas cuestiones se encuentra nuestra capacidad de ver la direccionalidad del comportamiento en el mundo, lo cual subyace a la idea de intención[38]. A su vez, estados mentales de mayor complejidad suponen la adquisición y entrenamiento de diversas habilidades, con lo que se van incorporando cuestiones propias del desarrollo de un lenguaje o de la posibilidad de especular acerca de los

---

36 Con estos argumentos Wittgenstein también rechaza lecturas conductistas que reducen lo mental a conductas observables (junto a propiedades físicas de los mismos), pues, por una parte, en lo mental hay un elemento de sentido que no puede ser reducido a ello y, por otra, gran parte de nuestra vida mental se juega en cuestiones que van más allá de experiencias corporales (véase Budd, 1993: 10–15; Hacker, 1997: 4–5, Moya, 2004: 62–76). En otro lugar he propuesto interpretar a esta perspectiva como disposicional-normativa (véase Figueroa Rubio, 2015; 2019: cap. III.2.2, IV.2.1). Esta perspectiva es, en principio, compatible con diversas tesis defendidas por funcionalistas, conexionistas, y defensores de teorías de la mente extendida y de la mente corporalizada. Dichas semejanzas y diferencias no serán revisados en este texto.

37 Sobre los reportes provenientes de la primera persona, Wittgenstein señala: "'Creo…' ilumina mi estado. De esta expresión se pueden inferir conclusiones acerca de mi conducta. Por tanto, aquí hay un parecido con las expresiones de los cambios emocionales, de la disposición, etcétera" (Wittgenstein, 1953: 260).

38 Esto se refleja en que "en la descripción del contenido de una intención, utilizamos las mismas palabras que utilizamos para señalar qué cuenta como la satisfacción de la intención" (Prades, 2011: 73).

estados mentales de otros con consecuencias tales como generar acciones que pueden identificarse como mentir o simular. En esta línea, en la sección 362 de *Investigaciones filosóficas* se señala:

> Un niño tiene que aprender mucho antes de poder disimular. (Un perro no puede ser hipócrita, pero tampoco puede ser sincero).

Y en el volumen I de *Últimos escritos sobre filosofía de la psicología* se lee:

> Solo en un complicado *juego de expresiones* puede haber simulación y su opuesto. (Igual que un movimiento falso o correcto se da en un juego) (WITTGENSTEIN, 1982: 208-208).

Un rasgo de varios estados mentales es que pueden ser simulados, lo cual requiere del desarrollo de capacidades más complejas y, con ello, diversos tipos de juicios en relaciones de segunda persona y tercera persona con otras que permiten una mayor complejidad en las interacciones. El espacio lógico de lo animal es un escenario complejo en donde las diferentes habilidades están distribuidas de forma variada entre los individuos de las diversas especies[39].

En conclusión, cuando hablamos de lo mental no referimos meramente a estados cerebrales, más bien, predicamos cuestiones psicológicas de una criatura viva que caza, tiene hambre, penas y que normalmente manifiesta su sentir con su comportamiento; precisamente es eso lo que hace que sea su comportamiento y que otros individuos puedan percibirlo como tal (véase HACKER, 1997: 46)[40]. La relación entre la mente y su expresión es interna. Así, pensando en la forma en que se puede entender el comportamiento de un joven como delincuente, Narváez señala:

> Conocer la sinestesia o el caso del cerebro del adolescente supone hacerse cargo de una explicación, no de una descripción que vincula elementos conectados

---

39 Este punto es tratado por Wittgenstein al preguntarse si tiene sentido atribuir estados mentales complejos a perros (véase WITTGENSTEIN, 1953: § 250, § 650).

40 Puede afirmarse que "cuando se ve la conducta del ser vivo, se ve su alma" (WITTGENSTEIN, 1953: § 357).

> lógicamente. Si decimos que ahora entendemos cómo se produce la conducta del adolescente, cometemos un error al equiparar la comprensión de qué es (en qué consiste ser) conducta de adolescente y cómo están los cerebros de los adolescentes con esa conducta (Narváez, 2016: 117).

Frases como "me gusta esa canción" manifiestan dicho gusto y no simplemente describen algo que sucede en el *interior*, y lo mismo sucede con poner una cara de alegría al escuchar la canción y ponerse a bailar con ella. Ambas cuestiones sirven como criterios para adscribir el estado mental a una persona; aunque dichas adscripciones sean derrotables, en principio es la forma en que se vive la vida mental. Como señala Finkelstein, si una persona desea saber mi condición psicológica, usualmente yo soy la mejor persona a quien preguntar por la misma razón que mi cara es el mejor lugar para buscar la expresión de dichos estados mentales, ambas son expresiones de lo que sucede en mí que requieren de diversas capacidades y tienen diversas características (véase Finkelstein, 2008: 101). En palabras de Anthony Kenny: "la expresión física de un proceso mental es un criterio para ese proceso; es decir, el que un proceso mental de un tipo particular deba tener una manifestación característica es parte del concepto de ese proceso" (Kenny, 1989: 33)[41]. Los movimientos del cuerpo humano no son (solo) contracciones de músculos, sino que son sonrisas, caras de asco, voces de enojo, miradas de desprecio.

La visión defendida en estas páginas nos lleva a entender que nuestra comprensión de lo mental está fuertemente vinculada a una comprensión del comportamiento, aunque se pueden tener instancias de estados mentales que no se expresan en comportamientos, o personas que, por diversas cuestiones, no están en condiciones de realizar los comportamientos relevantes. Creo que esto es clave para la comprensión del fenómeno jurídico. A ello me referiré brevemente en lo que sigue.

---

41 En esta misma línea, utilizando la distinción wittgensteiniana entre criterio y síntoma, Georg von Wright señala que los estados neuronales pueden ser considerados como síntomas, pero no como criterios para la identificación y atribución de estados mentales (véase von Wright, 1985: 108-115; 1997: 110-112; González Lagier, 2012: 507-509; Hacker, 1997: 45-51).

5. Expresión y adscripción de acciones

Retomando algunas de las cuestiones dichas, podemos ver que el rol que juegan las interacciones con otros muestra que la mente no es una cosa. Pardo y Patterson, inspirándose en Aristóteles, la comprenden como un conjunto de poderes y habilidades que se predican de los individuos, en cuanto complejos, no de parte de su cuerpo. Lo mental es así visto como un conjunto de habilidades, capacidades y poderes que poseen los humanos y otros animales en diferentes medidas. Cuando se identifican en casos concretos se atribuyen al animal completo, no a su cerebro. Es el gato el que caza, no el cerebro del gato, es el niño el que está feliz, no el cerebro del niño. A su vez, esto permite ver que en nuestras prácticas cotidianas nos enfrentamos a otros por medio de este tipo de atribuciones.

Estas atribuciones pueden ser puestas en duda y ser fruto de negociación e interpretación, como se ha señalado. Esto no deja de ser relevante, pues permite dar cuenta de cómo parte de nuestra identidad se construye por medio de nuestro comportamiento y su significación para otros (véase Coulter, 1979: cap. 2; Shotter, 1993: cap. 1; Medina, 2009: cap. 3). Así, por ejemplo, se puede ser conocido como alguien de buenas intenciones o como alguien mentiroso. A su vez, se puede ser conocido como un asesino pérfido o como alguien que mató a una persona por accidente, dependiendo de cómo se lea lo ocurrido en el contexto. De esta forma, se suele entender que nuestras acciones hablan por nosotros y manifiestan quiénes somos. Esto es relevante para el ámbito jurídico, pues su foco de atención suelen ser las acciones de las personas, ya sea para identificarlas como delitos, ya sea para asumir que manifiestan la aceptación de una norma, ya sea para asignarles efectos jurídicos, entre otras cosas. Teniendo en cuenta esto, lo dicho permite clarificar algunas cuestiones en torno a la atribución de acciones. Quisiera dedicar lo que queda de este trabajo a realizar una breve reflexión acerca de ello.

El problema que Finkelstein y Stroud detectan al aplicar el análisis del seguimiento de reglas a la atribución de estados mentales tiene su símil en el caso de pensar la acción. En el caso de la acción, la filosofía contemporánea ha trabajado en gran medida bajo el supuesto de que lo que marca la diferencia entre lo que hacemos y lo que nos ocurre se encuentra en la voluntad o el querer. Bajo este modelo, las acciones son movimientos

corporales a los que se suma la voluntad o querer. De ahí que gran parte de la discusión se centre en cómo entender dicho elemento.

Ludwig Wittgenstein, en la sección 621 de *Investigaciones filosóficas*, caracteriza la pregunta en que se centra dicho modelo de la siguiente manera:

> Pero no olvidemos una cosa: cuando 'levanto mi brazo', se levanta mi brazo. Y el problema se produce: ¿qué es lo que queda, cuando del hecho de que levante mi brazo sustraigo el que mi brazo se levante?

En las observaciones en torno a esta pregunta, Wittgenstein revisa críticamente dos posibles caminos que se pueden seguir[42]. Según el primero, la voluntad o el querer sería algo que experimentamos. Cada acción sería algo que nos sucede y que, con el tiempo, aprendemos a reconocer y controlar. El problema de esta idea es que niega que las acciones sean algo que hacemos, cuestión que está directamente en el corazón de la idea de acción y agencia: las acciones son precisamente aquello que hacemos y no meramente algo que nos sucede. El segundo camino supone asumir que sí somos quienes producimos nuestros movimientos corporales y que lo hacemos a través de un estado mental específico; acá se suele hablar de voliciones, actitudes favorables o intentos. Lo extraño de este camino es que nos sitúa como observadores de nuestra propia acción, además de que surge el problema de tener que responder qué es aquello que produce la volición (¿otra volición?), lo que nos conduce al problema del regreso al infinito al que se hizo referencia unas páginas atrás. Podemos ver que ambos caminos tienen su símil en las propuestas de la filosofía de la mente revisadas en la sección III: o bien mi cerebro provoca mis acciones, cuestión de la que después me doy cuenta, o bien yo domino mi cerebro y provoco mis acciones. De todas formas, esto nos conduce a un callejón sin salida, pues aquellos movimientos corporales a los que llamamos acción no parecen ser producidos por nosotros, pero tampoco tiene sentido que no sean producidos por nosotros[43].

---

42 He analizado estas observaciones en Figueroa Rubio, 2016: 147-158. Una presentación crítica en Hyman, 2015: 13-20.

43 Una consecuencia de esta última posición es tener que renunciar a la idea de agencia voluntaria, al asumir que somos producto de nuestro cerebro. Pero, tal como sucede con

Pero el problema está en la pregunta misma, en suponer que hay *algo* que es distinto del cuerpo y que puedo conocer, ya sea porque aprendo a reconocerlo por medio de mis experiencias o por ser un estado específico que produzco, que autoriza a decir que ciertos movimientos corporales son voluntarios y otros no. En este caso, la pregunta presente en la citada sección 621 supone un abismo entre levantar mi brazo y *lo que queda* además del hecho de que el brazo se levante.

En este ámbito también puede presentarse una alternativa a esta dicotomía que surge del dualismo movimiento corporal/querer. Dualismo que sufre de un problema similar al que afecta la distinción entre unas manchas de tinta sobre un papel y su significado. El texto de la sección 615 de *Investigaciones filosóficas* expresa:

> "La volición, si no se considera una especie de deseo, tiene que ser la acción misma. No puede detenerse antes de la acción". Si es la acción, entonces lo es en el sentido usual de la palabra; o sea, hablar, escribir, andar, levantar algo, representarse algo. Pero también: tratar, intentar, andar, levantar algo, representarse algo, etcétera.

---

la reducción de lo mental a lo cerebral, este tipo de reduccionismo deriva en sinsentido. Esto lo muestra Maribel Narváez al comparar la idea de que un no creyente derriba un mito (*e.g.*, la existencia de Dios) a favor de la verdad con la idea de que la agencia voluntaria es un mito. Señala: "el problema de la acción intencional no puede zanjarse del mismo modo que la cuestión de ser creyente. Esto es, aquí no podemos decir que hay dos prácticas alternativas: una para la cual la creencia en Dios va en contra de toda evidencia científica, que dice del creyente que es un dogmático aferrado a una ilusión, y otra, la del creyente, que llama al ateo materialista, que solo se guía por un criterio para seleccionar su participación en prácticas, el de la verdad. Sabemos que es posible vivir sin dios, de mejor o peor manera, pero no podemos ni tan solo dotar de sentido a una práctica en la que no se es agente (intencional y responsable): ¿habría alguien ahora realizando la acción intencional de escribir estas líneas? ¿Es ese hecho un acaecimiento?" (Narváez, 2016: 115).

Por otra parte, también se presenta el problema de no poder dar cuenta de cómo la voluntad produce las acciones, tal como surgía el problema en filosofía de la mente (véase Moya, 1990: cap. 1-2).

La respuesta de Wittgenstein en este punto es similar a la presentada unas páginas más arriba y se complementa con lo señalado respecto de la atribución de estados mentales. No hay que buscar algo fuera del comportamiento contextualizado para saber si una persona hace algo o qué es lo que está haciendo. En consecuencia, también podemos hablar de la acción como expresión, en contraste con teorías que, en un sentido amplio, se denominan causalistas[44].

Charles Taylor, en su ensayo *Acción como expresión* (1979), presenta una forma de entender a la acción desde una lectura expresivista. En primer lugar, señala que la forma en que una acción se expresa es directa, no se apoya en una inferencia. A diferencia de otras formas no inferenciales de acceder a un objeto, lo característico de una expresión es que "lo que se expresa solo puede ser manifiesto en la expresión" y que "el objeto expresivo revela lo expresado en un sentido más fuerte que el mero hecho de permitir verlo" (Taylor, 1979: 79). En resumen, según Taylor:

> Lo esencial aquí es que concebimos al objeto expresivo no solo como si nos permitiera ver algo, sino, en cierto sentido, como si nos dijera algo. […] [E]n el caso de los objetos expresivos, lo que expresan/dicen/manifiestan es algo que, en cierto modo, *hacen*, y no algo que pueda ocurrir por su intermedio. Y por esa razón, lo que expresa una cosa no puede reducirse a la posible lectura hecha en ella por un observador (Taylor, 1979: 80)[45].

Si bien lo expresado no es idéntico a la expresión, la relación que hay entre lo expresado y la expresión no puede ser reducida a una relación como la de causalidad en la cual tenemos dos cuestiones distintas donde una es efecto de la otra.

De este modo, la mente de otros se nos presenta en lo que otros hacen ante nosotros. En este punto la actividad corporal juega un rol muy importante, pues es a través de dicha actividad que se hace presente en el

---

44 Sobre las teorías causalistas de la acción y sus diversas versiones, véase Aguilar y Buckareff, 2010. Sobre sus efectos en el derecho, con especial énfasis en el derecho penal, véase Hart, 1960; Moore, 2010.

45 Y véase Schatzki, 1996: 31-34.

mundo lo que sucede a una persona (véase SCHATZKI, 1996: 21–25, 41–54)[46]. A su vez es fuente de conocimiento y especulación para otros, quienes pueden acceder a lo que sucede en otros y, a su vez, reaccionar ante ello. Esto constituye un fluir continuo de interacciones en las cuales la actividad corporal adquiere sentido dentro de una historia.

La interacción con otros es fundamental porque es en dicha interacción en la que aprendemos a expresar y a atribuir estados mentales y, con ello, a dominar conceptos mentales. Esto, a su vez, muestra que el cuerpo no solo se define por su materialidad y sus aspectos mecánicos, sino que el cuerpo pasa a ser *significado.* Un cuerpo inmóvil en un determinado contexto puede comunicarnos lo que sucede en una persona de la misma forma en que lo hace por medio de conductas determinadas; lo que dicho cuerpo comunique se define a partir de la forma en que se debe leer en el contexto lo que ocurre con esa persona[47].

Si consideramos el contexto de atribución de acciones, debemos tener en cuenta que nuestro comportamiento es expresivo de nuestras intenciones y que, por defecto, se realiza voluntariamente. Como se señaló, no es por inducción que vinculamos mente y conducta. En nuestro día a día no estamos esforzándonos en provocar nuestras acciones, pero tampoco son cuestiones que nos ocurran. A su vez, nos enfrentamos a los demás como si su comportamiento fuera voluntario y expresivo de sus estados mentales, y no buscando aquello que está detrás de sus movimientos corporales para entender que actúan voluntariamente. En este espíritu, en la sección 577 de *Zettel* (1967) podemos leer:

> Son voluntarios ciertos movimientos con su ámbito normal de propósito, aprendizaje, intento, acción. Aquellos movimientos de los que tiene sentido

---

46 Como señala este autor, esto no significa negar las experiencias personales que tienen los individuos acerca de sus estados mentales, ni reducir dichos estados a su expresión corporal, pero son estas cuestiones las que son de nuestro principal interés en el contexto de nuestra discusión.

47 Considerando esto, no debemos confundir voluntariedad con movimientos corporales, ni poner como punto de partida de una acción la identificación de ciertos eventos entendidos como descripción de movimientos corporales (véase HYMAN, 2015: 25-66).

> decir que a veces son voluntarios y a veces involuntarios son movimientos pertenecientes a un ámbito especial.

Nos enfrentamos a los otros como si se comportaran voluntariamente, en contextos anormales surge la pregunta acerca de la voluntariedad e involuntariedad de su conducta, de otro modo sería imposible el intercambio con otros. Como argumento final de este trabajo, quisiera llamar la atención respecto de cómo la comprensión acerca del modo en que nos enfrentamos al comportamiento de otros fue desarrollada en el ámbito de la filosofía de la acción por los autores denominados adscriptivistas[48]. Así, por ejemplo, John Austin señala en *A plea for excuses:*

> Me siento en mi silla del modo normal –no estoy embotado o influido por amenazas o cosas parecidas: en este caso, no valdrá decir ni que me senté en ella intencionadamente ni que no me senté en ella intencionadamente, ni tampoco que me senté en ella automáticamente o por hábito o lo que se quiera (Austin, 1956: 181).

Pero, a partir de esto, el adscriptivismo incorpora un par de argumentos más acerca de lo que implica atribuir acciones. Una idea central, que le da el nombre a la propuesta, es que las oraciones por medio de las cuales se habla habitualmente de acciones, tales como "él lo hizo" o "ella camina", no buscan describir la conducta de alguien, sino atribuírsela, de forma análoga a como oraciones como "esto es de él" o "ella tiene el derecho a disponer de ese bien" atribuyen propiedad y derechos (véase Hart, 1949). Considerando lo antes dicho, según esta tesis adscriptivista, las acciones se predican de las personas, no de sus movimientos corporales o de parte de su cuerpo.

---

48 Los autores adscriptivistas también han identificado en Aristóteles el antecedente de las tesis defendidas. Esto podría hacernos pensar que asumir este tipo de teoría supone volver a una teoría dualista de lo mental. Al respecto hay que decir dos cosas. La primera es que el adscriptivismo es una teoría de la acción, antes que de lo mental; la segunda es que el dualismo asumido por Aristóteles es diferente al defendido por Descartes, y es plausible pensar que no defiende la estructura argumental dualista antes referida (véase Pineda, 2012: cap. 2).

Por otra parte, para el adscriptivismo, la identificación de una acción como de un determinado tipo depende de la forma en que se le asigna sentido dentro de una comunidad concreta, interpretándola dentro de un contexto comunitario amplio que supone la participación en una historia de interacciones en las que se aprende a identificar patrones, cambios y reacciones en el comportamiento de otros a partir de parámetros comunes. En esta línea, Theodore Schatzki identifica al menos cuatro elementos del contexto que son relevantes para poder atribuir una acción y, con ello, un estado mental específico: la conducta pasada y futura de la persona a la que se atribuye, la red de condiciones de vida que forman parte de lo que dicha persona es (*e.g.*, sus capacidades intelectuales o físicas), la situación en la cual la persona actúa y las prácticas en las que participa (SCHATZKI, 1995: 35-36, 55-87). A ello hay que incorporar elementos de la interacción misma en que se atribuyen dichos estados mentales, pues muchas veces en algunos contextos deben seguirse reglas específicas para atribuir correctamente acciones y estados mentales que en otros se ignoran, por lo que la perspectiva de quien atribuye tiene un rol especial. Todos estos elementos están, a su vez, marcados por un trasfondo de representaciones comunes que se transmiten entre personas y que se presentan como condiciones de una interacción exitosa. Una correcta atribución de un estado mental, así como la correcta identificación de un comportamiento como de un tipo determinado, se define dentro de estas variables, y un estudio de este tipo de juicios de atribución debe tenerlos en cuenta.

CONSIDERACIONES FINALES

Me parece que con una lectura que tenga presentes tanto ideas adscriptivistas como expresivistas sobre la acción se puede dar cuenta de la realidad jurídica de una forma que haga justicia a la práctica jurídica y que evite varios de los problemas que hemos visto a lo largo de este trabajo[49]. En los discursos jurídicos usualmente se considera que existe una interacción

49 Cabe señalar que una lectura expresivista de la acción permite evitar los problemas que identifica Federico Arena (ARENA, 2008) en el adscriptivismo, pues no la hacen depender de una comprensión fisicalista de lo mental.

entre mente y acción, y la atribución de una acción es la atribución de una intención y dicha atribución se hace a una persona, no a una parte de su cuerpo. A su vez, cuando pensamos la vía por medio de la cual accedemos a ello, podemos darnos cuenta de que es la de los comportamientos (o su ausencia), los que, a su vez, normalmente asumimos como voluntarios y como expresivos de la intención en cuestión. Así, por ejemplo, al acusar que un ilícito se ha cometido con culpa y no con dolo apelamos a criterios conceptuales que permiten entender la conducta de los individuos en su contexto como expresivas de descuidos, no tomarse en serio las cosas, tener malas intenciones y hacer planes, entre otras múltiples. Esto es relevante porque las preguntas acera de qué diferencia una conducta dolosa de una negligente es una pregunta conceptual que no puede ser respondida por medio de estudios empíricos. Una disputa acerca de la atribución de cualquiera de esos estados a una persona se realiza situándola en un contexto complejo y no mediante la identificación de la activación de ciertas zonas de su cerebro.

Lo anterior muestra que es necesario diferenciar lo conceptual de lo empírico a la hora de hablar de lo mental y, a su vez, ser cautos a la hora de pensar la reducción de lo uno a lo otro. Algo similar puede decirse sobre lo que implica aceptar una norma, concepto mental con el que la filosofía del derecho ha trabajado largamente en las últimas décadas. Identificar el compromiso de un individuo con una norma o con un sistema de normas es una cuestión compleja que supone preguntarse qué es lo que hace que podamos atribuir dicho compromiso a alguien, lo que nos lleva a formular preguntas como: ¿supone la defensa de razones morales o se puede hacer solo a partir de razones prudenciales?, y ¿supone la posibilidad de no aplicar la norma para resolver casos difíciles? Las cuestiones en juego al momento de atribuir la aceptación de una norma no se pueden reducir a la lectura de una neuroimagen sin dejar de hablar de lo que es relevante en aquel concepto mental.

El blanco de ataque de este texto no es la existencia de la neurociencia cognitiva, ni su importancia para saber cómo nuestro sistema nervioso funciona y, con ello, aprender de nuestro comportamiento. Por el contrario, el valor de los avances en esta ciencia para conocernos es difícil de exagerar y su utilización en los procedimientos judiciales debe ser objeto de una discusión distinta a la aquí expuesta. Por otra parte, en estas páginas no

se ha puesto en duda que la presencia de un sistema nervioso funcionando normalmente en un individuo (en tanto miembro de una especie) es una condición necesaria para que puedan realizarse cierto tipo de conductas y contar con ciertos estados mentales. Además, en este trabajo no se ha negado que cambios en el sistema nervioso puedan significar cambios en la conducta, ni que el conocimiento que de esto podamos obtener por medios científicos pueda afectar nuestros juicios sobre las personas (*e.g.*, declarándolas inimputables en un procedimiento penal). Tampoco es necesario negar que cada vez que una persona tiene un estado mental, existe un estado cerebral concreto con el cual sea posible establecer una correlación.

Los tres pasos que han sido objeto de reflexión dicen algo distinto de las cuestiones presentes en el párrafo anterior, pues procuran reducir nuestra comprensión de lo mental y de la agencia a lo cerebral, y ello supone abrazar una serie de supuestos que tienen consecuencias, a mi parecer, indeseadas. Ahora, en estas páginas sí se ha llamado la atención respecto de que al hablar de lo mental es en el nivel personal e interpersonal donde los conceptos tienen sentido, no en el subpersonal, para utilizar el vocabulario desarrollado por Daniel Dennett (2013: cap. III)[50]. Esto se refleja, por ejemplo, en que, dependiendo del conocimiento con que contemos sobre el estado mental de una persona, trataremos a esa persona como imputable o inimputable, y no a su cerebro, o en que, cuando decimos que alguien ha tomado una decisión, decimos que ella lo ha hecho, no su cerebro. En el denominado neuroderecho es muy común confundir ambas cuestiones, diciendo que el cerebro de los jueces *adjudica* o que el dolo *es* un específico estado cerebral[51].

En estas páginas se ha defendido un modelo que propone asumir un continuo entre mente, comportamiento y acción para dar cuenta de la forma en que se atribuyen estados mentales y, con ello, se utilizan los conceptos mentales en el contexto jurídico. Especialmente cuando consideramos que el contexto jurídico es uno en el cual la agencia juega un rol central. A partir de esto no se niega la relevancia de los avances de la neurociencia, pero sí

---

50 Una discusión en torno a sus visiones sobre estas cuestiones en Bennett *et al.*, 2007.

51 Un conjunto de ensayos interesantes sobre los diversos alcances de estos discursos en áreas específicas del derecho en Freeman, 2011.

se acusa la posibilidad de caer en sinsentidos al adoptar irreflexivamente el argumento de tres pasos al que se ha hecho referencia[52]. Es necesario preguntarse por el sentido que tiene hablar de lo mental en el día a día, en los diversos contextos en que nos desenvolvemos, y de qué forma lo hacemos. Respecto de este tipo de preguntas, se ha propuesto considerar al comportamiento habitual de las personas como expresivo de lo mental, y que nuestros juicios sobre las acciones de otros son adscriptivos, en ellos aplicamos lo aprendido en comunidad, a partir de la identificación de patrones y estándares comunes. En consecuencia, es posible afirmar, contra el modelo que nos dice que debemos elegir entre dualismos (ya sea el dualismo de sustancias, o el fiscalista que distingue entre cerebro y conducta), que los humanos no son cuerpos portadores de una mente o cerebro, sino que, por el contrario, son sujetos que disfrutan, temen, delinquen y aceptan normas en su interacción con otros.

REFERENCIAS

Aguilar, J. y Buckareff, A., 2010: "The causal theory of action: origins and issues", en Aguilar, J. y Buckareff, A. (eds.), *Causing Human Actions.* Cambridge (ma): mit Press, 1–26.

Aguilar, J. y Buckareff, A. (eds.), 2010a: *Causing Human Actions.* Cambridge (ma): mit Press.

Anscombe, G. E. M., 1963: *Intention.* 2.ª ed. Cambridge (ma), London: Harvard University Press.

Arena, F., 2008: "La identificación de la acción individual", *Analisi e Diritto*: 203–227.

52 Esto no quiere decir que no cambie nuestra forma de utilizar estos conceptos con el desarrollo del conocimiento científico. No se trata de defender una postura rígida que asuma un quietismo o una impermeabilidad de nuestra comprensión del mundo y de nuestro uso del lenguaje (véase Figueroa Rubio, 2019a).

AUSTIN, J. L., 1956: *A Plea for Excuses.* Citado por la traducción al castellano "Un alegato en pro de las excusas", en AUSTIN, J. L., *Ensayos filosóficos.* 2.ª ed. Madrid: Alianza, 1989, 169-192.

AUSTIN, J. L., 1989: *Ensayos filosóficos.* 2.ª ed. Madrid: Alianza.

BENNETT, M. R. y HACKER P. M. S., 2005: *Philosophical Foundations of Neuroscience.* Hoboken (NJ): Wiley-Blackwell.

BENNETT, M. R. y HACKER, P. M. S., 2007: "Fragmentos de *Philosophical Foundations of Neuroscience*", en BENNETT, M. R., HACKER, P. M. S., DENNETT, D. y SEARLE, J., *La naturaleza de la conciencia*. Barcelona: Paidós, 15-68.

BENNETT, M. R., HACKER, P. M. S., DENNETT, D. y SEARLE, J., 2007: *La naturaleza de la conciencia*. Barcelona: Paidós.

BRONCANO, F. (ed.), 1995: *La mente humana. Enciclopedia Iberoamericana de Filosofía.* Vol. 8. Madrid: Trotta.

BROŻEK, B. y STELMACH, J. (eds.), 2017: *The Province of Jurisprudence Naturalised*, *Revus*, 32.

BUDD, M., 2014: *Wittgenstein's Philosophy of Psycholology.* Abingdon-on-Thames: Routledge.

CORBÍ, J. y PRADES, J. L., 2000: *Mind, Causes and Mechanisms.* Hoboken (NJ): Wiley-Blackewell.

COULTER, J., 1979: *The Social Construction of Mind.* London: The MacMillan Press.

CRANE, T., 1995. "Mental causation", *Proceedings of the Aristotelian Society*, 69: 211-253.

DARWALL, S., 2006: *The Second Person Standpoint.* Cambridge (MA): Harvard University Press.

DAVIDSON, D., 1980: *Essays on Actions and Events.* Oxford: Oxford University Press.

Dennett, D., 2013. *Intuition Pumps and Other Tools for Thinking.* New York (NY): W. W. Norton & Co.

Dummett, M., 1959: "Wittgenstein's Philosophy of Matemathics", *Philosophical Review,* 68 (3): 324-348.

Figueroa Rubio, S., 2015: "Expectativas y atribución de responsabilidad" *Revus,* 26: 93-110.

Figueroa Rubio, S., 2016: "Reconocer y explicar (,) lo que hacemos", en Ormeño, J., *Acciones, razones y agentes.* Santiago de Chile: LOM, 145-178.

Figueroa Rubio, S., 2019: *Adscripción y reacción. Responsabilidad jurídica y moral desde una perspectiva interpersonal.* Madrid: Marcial Pons.

Figueroa Rubio, S., 2019a: "Pena, determinismos y la imagen que tenemos de nosotros mismos", *Anuario de Filosofía Jurídica y Social,* 33: 37-58.

Finkelstein, D., 2003: *Expression and the Inner.* Cambridge (MA): Harvard University Press.

Freeman, M., 2011: *Law and Neuroscience.* Oxford: Oxford University Press.

Gabriel, M., 2016: *Yo no soy mi cerebro.* Barcelona: Ediciones Pasado y Presente.

González Lagier, D., 2012: "¿La tercera humillación? (Sobre neurociencia, filosofía y libre albedrío)", *Doxa. Cuadernos de Filosofía del Derecho,* 35: 499-510.

Hacker, P. M. S., 1997: *Wittgenstein on Human Nature.* UK: Phoenix.

Hart, H. L. A., 1949: *Ascription of Responsibility and Rights.* Citado por la traducción al castellano "La adscripción de responsabilidad y de derechos", *Revista de Teoría del Derecho,* 1 (1), 2014: 13-32.

Hart, H. L. A., 1960: "Acts of will and responsibility", reimp. en Hart, H. L. A., *Punishment and Responsibility.* Oxford: Oxford University Press, 90-11.

Hart, H. L. A., 1968: *Punishment and Responsibility.* Oxford: Oxford University Press.

Hyman, J., 2015: *Action, Knwoledge and Will.* Oxford: Oxford University Press.

Illes, J. y Sahakian, B. (eds.), 2011: *The Oxford Handbook of Neuroethics.* Oxford. Oxford University Press.

Kenny, A., 1989: *The Metaphysics of Mind.* Citado por la traducción al castellano *La metafísica de la mente.* Barcelona: Paidós, 2000.

Kim, J., 1996: *Philosophy of Mind.* Abingdon-on-Thames: Routledge.

Libet, B., 1985: "Unconscious cerebral initiative and the role of conscious will in voluntary action", *The Behavioral and Brain Sciences,* 8: 529-566.

Lowe, E. J. 2008: *Personal Agency. The Metaphysics of Mind and Action.* Oxford: Oxford University Press.

Medina, J., 2009: *Speaking from Elsewhere.* New York (ny): suny Press.

Miller, A. y Wright, C., 2002: *Rule-Following and Meaning.* Montreal: McGill-Queen's University Press.

Moya, C., 1990: *The Philosophy of Action.* Cambridge: Polity Press.

Moya, C., 2006. *Filosofía de la mente.* 2.ª ed. Valencia: Universitat de Valencia.

Moore, M. S. 2010: "Renewed questions about the Causal Theory of Action", en Aguilar, J. y Buckareff, A., *Causing Human Actions.* Cambridge (ma): mit Press, 27-44.

Narváez, M., 2015: "Expresiones de normas. Sobre formulaciones normativas y otras entidades en la Teoría del Derecho", *Revus*, 25: 71-100.

Narváez, M., 2016: "Neurodisciplinas y acción: el sentido como límite", en Ormeño, J., *Acciones, razones y agentes.* Santiago de Chile: lom, 97-122.

Ormeño, J. (ed.), 2016a: *Acciones, razones y agentes.* Santiago de Chile: lom.

Pardo, M. y Patterson, D., 2013: *Minds, Brains, and Law.* Oxford: Oxford University Press.

Prades, J. L., 2011: "Why enough should be enough: The role of expressive behaviour", *Teorema: International Journal of Philosophy,* 30 (3): 67-78.

Pérez, D., 2013: *Sentir, desear, creer.* Buenos Aires. Editorial Prometeo.

Pineda, D., 2012: *La mente humana.* Madrid: Cátedra.

Ramsey, W., 2019: "Eliminative Materialism", en Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy (Spring 2019 Edition).* Disponible en: https://plato.stanford.edu/archives/spr2019/entries/materialism-eliminative/ (visitada en abril de 2019).

Reiner, P., 2011: "The rise of Neuroessentialism", en Illes, J. y Sahakian, B. (eds.), *The Oxford Handbook of Neuroethics.* Oxford: Oxford University Press, 161-175.

Rubia, F., 2009: "Comentarios introductorios", en Rubia, F., *El cerebro: avances recientes en neurociencias.* Madrid: Editorial Complutense, 15-20.

Rubia, F. (ed.), 2009a: *El cerebro: avances recientes en neurociencias.* Madrid: Editorial Complutense.

Rubia, F., 2015: *El cerebro espiritual.* Barcelona: Fragmenta.

Rudder Baker, L., 1995: *Explaining Attitudes. A Practical Approach to the Mind.* Cambridge: Cambridge University Press.

Ryle, G., 1949: *The Concept of Mind.* Chicago: University of Chicago Press.

Schatzki, T., 1996: *Social Practices. A Wittgensteinian Approach to Human Activity and the Social.* Cambridge: Cambridge University Press.

Schultz, W., 2015: "Neuroessentialism: Theoretical and clinical considerations", *Journal of Humanistic Psychology* 58 (6): 1-33.

Shotter, J., 1993: *Conversational Realities.* London: Sage.

Stroud, B., 1965: "Wittgenstein and logical necessity", en Stroud, B., *Meaning, Understanding and Practice.* Oxford: Oxford University Press, 1-16.

STROUD, B., 1996. "Mind, meaning, and practice", en STROUD, B., *Meaning, Understanding and Practice*. Oxford: Oxford University Press, 171-192.

STROUD, B., 2000: *Meaning, Understanding and Practice*. Oxford: Oxford University Press.

TAYLOR, C., 1979: "Action as expression". Citado por la traducción al castellano "Acción como expresión", en TAYLOR, C., *La libertad de los modernos*. Buenos Aires: Amorrortu, 77-96.

TAYLOR, C., 2005: *La libertad de los modernos*. Buenos Aires: Amorrortu.

TIEHEN, J., 2018: "Physicalism", *Analysis,* 78 (3): 537-551.

TORIBIO, J., 1995. "Eliminativismo y el futuro de la psicología popular", en BRONCANO, F. (ed.), *La mente humana. Enciclopedia Iberoamericana de Filosofía*. Vol. 8. Madrid: Trotta, 245-272.

VON WRIGHT, G. H., 1985: *Sobre la libertad humana*. Barcelona: Paidós.

VON WRIGHT, G. H., 1997: "Notes on the Philosophy of Mind", en VON WRIGHT, G. H., *In the Shadow of Descartes. Essays in the Philosophy of Mind*. Dordrecht: Springer, 97-123.

VON WRIGHT, G. H., 1997a: *In the Shadow of Descartes. Essays in the Philosophy of Mind*. Dordrecht: Springer.

WEGNER, D., 2003: "The mind's best trick: How we experience conscious will", *Trends in Cognitive Science*, 7: 65-69.

WILLIAMS, M., 1999: *Wittgenstein, Mind and Meaning*. UK: Routledge.

WITTGENSTEIN, L., 1953: *Philosophische Untersuchungen*. Citado por la traducción al castellano de la 4.ª ed. de 2009, *Investigaciones filosóficas*. Madrid: Trotta, y por la traducción al inglés de la 4.ª ed. de 2009, *Philosophical Investigations*. UK: Blackwell Publishing.

WITTGENSTEIN, L., 1958: *The Blue and Brown Books*. Citado por la traducción al castellano *Los cuadernos azul y marrón*. Madrid: Tecnos, 1968.

WITTGENSTEIN, L., 1967: *Zettel*. Citado por la traducción al castellano *Zettel*. México: UNAM, 1979.

WITTGENSTEIN, L., 1982: *Letzte Schriften über die Philosophie der Psychologie,* t. I, *Vorstudien zum zweiten Teil der "Philosophische Untersuchungen*". Citado por la traducción al castellano *Últimos escritos sobre filosofía de la psicología*. Madrid: Tecnos, 1994.

III.
Acción intencional de interpretación judicial. Casos normales y sorpresivos

DIEGO MORENO CRUZ*

INTRODUCCIÓN

A partir de un aparato de análisis conceptual psicologista de la interpretación judicial, mi propósito en este capítulo es cuestionar la popular idea de que hay casos –comúnmente conocidos como los casos fáciles– respecto de los que la interpretación y aplicación del derecho por parte de las jueces no involucra el ejercicio de su discrecionalidad.

En primer lugar, expondré la perspectiva de análisis conceptual psicologista de la "interpretación judicial" (1). Una lente analítica que enfoca a la intención formada en la mente de la juez de proferir un acto de habla interpretativo (1.1), en relación con casos individuales normales o sorpresivos (1.2), casos que son diferenciados a partir de una representación intuitiva de la (a)normalidad por parte de la juez (1.2.1). En segundo lugar, añadiré una distinción metodológica entre dos tipos de dirección de análisis psicologista de la interpretación judicial (2): un análisis prospectivo y un análisis retrospectivo (2.1), y, sobre la base del aparato de análisis conceptual ofrecido en primer lugar, sostendré que las jueces ejercen discrecionalidad al proferir actos interpretativos en relación con casos individuales normales o sorpresivos (2.1.1). Por último, formularé algunas consideraciones finales.

1. ANÁLISIS PSICOLÓGICO DE LA INTERPRETACIÓN JUDICIAL

Con base en experimentos que consisten en problemas hipotéticos de elección práctica puestos a grupos de personas, la psicología cognitiva ha formulado hipótesis acerca de la influencia que los mecanismos mentales inconscientes tienen en las acciones que elegimos realizar en ocasiones diferentes. Además, que estas elecciones están moldeadas tanto por ideales normativos que aceptamos conscientemente, como por tendencias

---

* Doctor europeo en Filosofía del Derecho de la Universidad de Génova (Italia). Docente investigador del Centro de Investigación en Filosofía y Derecho (CIFD) y profesor de Filosofía del Derecho de la Universidad Externado de Colombia y del *Master Global Rule of Law and Constitutional Democracy* de las universidades de Génova (Italia) y Girona (España).

psicológicas o mecanismos inconscientes de nuestra mente. Y bien, una tendencia actual de los estudios acerca del razonamiento jurídico ha sido la de prestar atención a estas hipótesis en aras de aplicarlas a la descripción, explicación y comprensión de cómo funciona la mente de las jueces, o de las autoridades normativas en general (funcionarios), cuando interpretan y aplican el derecho.

En este capítulo ofreceré un aparato de análisis conceptual psicológico de la interpretación judicial, es decir, una representación teórica de la atribución de un significado a una disposición jurídica construida desde la perspectiva interna o psicológica de la juez. Un análisis que toma como centro de un mapa psicológico explicativo de la interpretación judicial a la intención formada en la mente de la juez de proferir un acto de habla interpretativo, con el que ella expresa la norma jurídica general aplicable a un caso individual, para solucionar correctamente una disputa o cuestión concreta; un análisis de la intención de proferir actos de habla interpretativos sobre la base del "conjunto motivacional subjetivo"[1] de la juez: por una parte, una red interconectada de estados mentales intencionales ("*network*") –creencias, deseos, expectativas y emociones– que informan el proceso de formación del estado mental "intención", y, por otra parte, la capacidad, habilidad, disposición, tendencias, hábitos y otros sucesos y estructuras o mecanismos causales, o no-intencionales, de la mente ("*background*") (Searle, 2010: 31). De acuerdo con este esquema de representación teórica de la mente, el *background* es un presupuesto de la formación y funcionamiento de los estados mentales intencionales y, también, de poder hacer cosas con las palabras (Searle, 2011: 119-128): por ejemplo, interpretaciones judiciales.

En pocas palabras, a la luz de esta perspectiva, analizaré la intención de la juez de proferir un acto de habla interpretativo para expresar la norma jurídica general aplicable al caso individual bajo examen. Un acto interpretativo que es proferido con la mira puesta en obtener un resultado que

---

1 Williams, 1981: 105: un conjunto de elementos que provee razones internas para la acción "*as dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent*".

está representado en la intención de la juez de proferirlo, a saber, dictar un veredicto o *decisum* correcto respecto a una disputa o cuestión jurídica concreta conocida por ella.

A continuación, indicaré una ambigüedad de la que adolece la "interpretación judicial" (1.1), y una distinción entre casos normales y sorpresivos, que es dependiente de los estímulos que los hechos del caso individual evocan en la mente de la juez (1.2).

1.1. Interpretación-actividad; interpretación-producto; interpretación-objeto

Por "interpretación judicial" se puede entender por lo menos tres cosas distintas: (*i*) interpretación-actividad; (*ii*) interpretación-producto; (*iii*) interpretación-objeto.

(*i*) *Interpretación-actividad*: proceso que se activa (in)conscientemente en la mente de la juez con la finalidad práctica de formar la intención de atribuir un significado a una disposición jurídica mediante un acto de habla interpretativo.

(*ii*) *Interpretación-producto*: la intención formada en la mente de la juez de proferir un acto de habla interpretativo usando un enunciado interpretativo –de la forma "la disposición jurídica *dj* significa la norma jurídica *nj*", o, en otros términos, "*nj* es la traducción correcta de *dj*"– como la norma jurídica general aplicable a un caso individual. Dicho de otro modo, la intención formada en la mente de la juez de proferir el acto interpretativo[2].

(*iii*) *Interpretación-objeto*: precisamente, el acto de habla interpretativo que es objeto de la intención de proferirlo.

Respecto a estas distinciones es pertinente hacer dos aclaraciones.

Por un lado, la interpretación-producto y la interpretación-objeto son acciones intencionales de naturaleza diferente. El estado mental "intención" es formado mediante un acto mental de elección, que implica el cambio de

---

2 Toda acción presupone la elección (in)consciente de acciones particulares entre otras acciones. "Elegir" no presupone necesariamente una deliberación o actividad consciente. Con "elegir" se hace referencia simplemente a que seleccionar una acción implica no elegir otras acciones alternativas. Al respecto, ver Simon, 1945, cap. I.

un proceso de formación de la intención a un estado mental final, a saber, la *intención* formada en la mente de la juez de proferir el acto de habla interpretativo elegido. Y, como sucede respecto a toda acción intencional, el acto mental de elección es especificable, y también explicable, sobre la base de los estados mentales intencionales de la juez (su *Network*) y los sucesos y mecanismos mentales no-intencionales (su *Background*) que influencian la formación de la intención en su mente de proferir un acto interpretativo. Por otro lado, la interpretación-objeto es un acto de habla interpretativo que es proferido con la intención de producir un resultado, un cambio en el mundo externo, es decir, un cambio de un estado inicial de cosas a un estado de cosas final que es deseado, imaginado, esperado, por la juez.

Hechas estas aclaraciones, se puede distinguir, a su vez, entre dos tipos de interpretación-actividad, dos tipos de interpretación-producto y dos tipos de interpretación-objeto.

I.I.I. DOS TIPOS DE INTERPRETACIÓN-ACTIVIDAD

Interpretación-intuitiva: un mecanismo mental de formación de la intención que opera rápidamente, sin esfuerzo, control, ni atención por parte de la juez, y que finaliza con un acto de elección inconsciente de proferir intencionalmente cierto acto interpretativo.

La juez no se plantea un problema de decisión acerca de qué significado atribuir a cierta disposición jurídica. Ella atribuye automáticamente el significado *prima facie* sin advertir otras opciones interpretativas, es decir, el significado que corresponde a sus hábitos lingüísticos, por ejemplo, el significado textual u ordinario y/o el significado que corresponde al propósito de la prohibición, permisión u obligación de la clase de acciones a las que hace referencia la disposición[3]. La juez no es consciente de la formación de las razones que informan la intención de atribuir el significado *prima facie*. Y a su vez, ella no tiene duda acerca de que el significado *prima facie* sea el significado atribuible a la disposición jurídica relevante. Ella forma

3 Sobre el significado ordinario, claro, de las disposiciones entendido como su significado textual, y también como el significado que corresponde al propósito, véase HART, 1983.

su intención en una situación de certeza (Hampshire y Hart, 1958: 1-12; Moreno Cruz, 2017: 151-213). El proceso de formación de la intención de proferir un acto de habla interpretativo, con el que ella hace uso del significado *prima facie* como la traducción autoritativa de una disposición jurídica, ocurre automáticamente. Este mecanismo involucra la activación de lo que en psicología cognitiva se conoce como el Sistema 1 del aparato cognitivo de la mente del agente (Kahneman, 2011: 19-58).

Interpretación deliberativa o *razonamiento interpretativo*: un proceso mental de formación de la intención que ocurre lentamente, y que la juez experimenta conscientemente que está ocurriendo en su mente; un proceso reflexivo que finaliza con un acto de elección consciente (*i.e.*, acto mental de decisión) de proferir intencionalmente cierto acto interpretativo y no otro acto interpretativo opcional.

Este proceso se activa en situaciones de incertidumbre o de indecisión (Hampshire y Hart, 1958; Moreno Cruz, 2017); la juez advierte que hay opciones interpretativas de cierta disposición, es decir, hay actos interpretativos alternativos y, de este modo, ella reconoce conscientemente un problema práctico acerca de cuál de esos actos interpretativos proferir (Hampshire y Hart, 1958; Raz, 1978; Pettit, 2010: 252-258); se trata de un proceso de razonamiento o de deliberación en el que la juez sopesa las razones a favor y en contra de atribuir el significado *prima facie* o un significado diferente a cierta disposición jurídica[4]. Ella tiene que decidir si, *all things considered,* confirmar, corregir o reemplazar el significado *prima facie* atribuible intuitivamente a la disposición. Este proceso involucra la activación de lo que en psicología cognitiva se conoce como el Sistema 2 del aparato cognitivo de la mente del agente (Kahneman, 2011: 19-58).

1.1.2. Dos tipos de interpretación-producto

A cada uno de los tipos de interpretación-actividad le corresponde un tipo de interpretación-producto diferente: la intención-presente que es formada por un acto de elección inconsciente que es el producto de una

4 Véase Hampshire y Hart, 1958: 3: "*'Shall I do it or shall I not?'*".

interpretación-intuitiva; y la intención-anticipada que es formada con un acto de elección consciente que es el producto de un razonamiento de toma de decisión interpretativa.

Intención-presente o elección intuitiva. El producto de la interpretación-intuitiva es una intención que se forma con un acto mental de elección automático e intuitivo, y que tiene por objeto proferir inmediatamente un acto de habla interpretativo. Se trata de la intención que se forma en la mente de la juez simultánea e inmediatamente al momento en que ella profiere el acto de habla interpretativo objeto de tal intención[5]. Es decir, una intención que acompaña y que está presente durante la realización del acto interpretativo ("*intention-in-action*") (Searle, 2010: 33-37). El estado mental "intención" se forma con un mecanismo mental de elección intuitiva; se trata, en otros términos, de una "*unconscious decision*" (véase Raz, 1978: 133-135): es decir, una intención que se forma sin que haya lugar a una deliberación ni a un acto de decisión previo[6].

Intención-anticipada o decisión. El producto de la interpretación-deliberativa es una intención que se forma con un acto mental de elección consciente o de toma de decisión en sentido estricto, y que tiene por objeto proferir un acto de habla interpretativo en el futuro. Una intención que es el producto de un balanceo de las razones a favor y en contra de proferir un acto interpretativo y no otro acto opcional (Hampshire y Hart, 1958: 12; Raz, 1978: 133-135; Searle, 2010: 33 y ss.; Hart, 2013: 657). Hay un fragmento de tiempo entre el momento en que la juez forma su intención y aquel en el que ella actúa (véase Raz, 1978: 133-135). En otras palabras, la decisión es tomada previamente, o la intención es formada anticipadamente, al momento en que se profiere intencionalmente el acto interpretativo ("*prior intention*" [Searle, 2010: 33]; "plan" [Bratman, 1987]).

De modo que, en ciertas situaciones, la intención de proferir cierto acto de habla interpretativo se forma en la mente de la juez sin que medie un

---

5 Sobre intenciones formadas simultáneamente a la realización de una acción, véase Petit, 2010: 252-253.

6 Hampshire y Hart, 1958: 9, 12: "*very frequently we know what we are doing, and so are doing it intentionally, although this has never been preceded by any stage of prior deliberation or doubt* […] *not all intention is formed as a result of prior decision*".

proceso mental consciente, es decir, sin que sea necesario un razonamiento (o deliberación) para producirla. Y, en otras situaciones, por el contrario, tal intención sí es el producto de una deliberación consciente.

1.1.3. DOS TIPOS DE INTERPRETACIÓN-OBJETO

Vale precisar que, sobre la base de la perspectiva psicologista, un acto de habla interpretativo es especificado con base en el resultado que –se asume– está representado en la intención de la juez de proferirlo. Por ejemplo, obtener el veredicto correcto, o la solución correcta de la disputa o cuestión concreta, que se deduce de la aplicación de la norma jurídica general, que la juez expresa con su acto de habla interpretativo, al caso individual.

Por otra parte, vale distinguir entre resultado y consecuencia de los actos de habla interpretativos[7]. El resultado es la obtención del cambio de un estado inicial de cosas a un estado final de cosas en el mundo, que está representado en la intención de la juez de proferir cierto acto interpretativo. La consecuencia es un cambio que, si bien no está representado conscientemente en la intención de la juez de proferir cierto acto, su obtención sí es un presupuesto no-intencional de la intención de proferirlo. Hemos dicho que el resultado que la juez quiere obtener con su acto interpretativo es obtener un veredicto correcto. Ahora bien, para que tal resultado se obtenga, se requiere de la producción de una consecuencia, un cambio no-intencional en el mundo, a saber, el hecho de que (en el futuro) otras jueces consideren su acto de habla interpretativo, y, por lo tanto, a su veredicto, "correcto"[8].

Una vez hechas estas aclaraciones, de las distinciones entre los dos tipos de actividades y productos interpretativos correspondientes, se puede distinguir entre dos tipos de interpretaciones-objeto o actos de habla interpretativos diferentes.

---

7 Sobre la distinción entre resultados y consecuencias de la acción intencional, véase VON WRIGHT, 1963.

8 Sobre la influencia de la creencia-expectativa de que jueces del futuro considerarán correcto el acto de habla interpretativo proferido por la juez en el presente, véase MORENO CRUZ, 2014: 237-251.

Actos interpretativos habituales. Actos con los que la juez hace uso de un enunciado interpretativo *prima facie* como la norma jurídica general aplicable a un caso individual.

Ahora bien, el uso del enunciado interpretativo *prima facie* puede ser objeto de una intención-presente, es decir, una intención que es el producto de una actividad intuitiva, o bien de una intención anticipada, es decir, una intención que es el producto de un razonamiento o de una deliberación: la decisión de confirmar el significado *prima facie* de la disposición como la norma jurídica general aplicable al caso. En ambas situaciones se puede afirmar que la juez actuó del modo habitual.

Actos interpretativos nuevos. Actos con los que la juez hace uso de un enunciado interpretativo diferente al enunciado *prima facie* como la norma jurídica general aplicable a un caso. Estos actos son objeto de una intención-anticipada, de una elección consciente, o de una decisión de proferirlos sobre otros actos interpretativos opcionales.

Una aclaración adicional: todo acto de habla interpretativo requiere de la formación de la intención-presente de proferirlo. Y para proferir en el futuro un acto interpretativo que es objeto de una elección consciente, de una decisión, se requiere de la intención-presente de hacerlo una vez llegado el momento planificado de proferirlo (Searle, 2010: 33-35).

1.2. CASOS NORMALES Y SORPRESIVOS

Que la interpretación judicial consista en un mecanismo mental intuitivo o en un razonamiento de toma de decisión interpretativo depende de los estímulos evocados en la mente de la juez al conocer los hechos que configuran el caso individual bajo examen.

Los estímulos evocados en la mente de la juez están condicionados por la inducción intuitiva e inconsciente de esperar, a partir de la experiencia acumulada en la memoria, que el futuro será similar al pasado. Una disposición innata de la mente de esperar que actos de habla interpretativos proferidos en ocasiones diferentes, en relación con casos individuales

similares, en circunstancias similares, tengan resultados y consecuencias similares[9].

Y bien, la inducción intuitiva es un presupuesto necesario del reconocimiento intuitivo de situaciones familiares y también de las situaciones que no son familiares. De ella depende que sea posible la producción de patrones de similitud perceptiva de los estímulos producidos por casos individuales en ocasiones diferentes, y la expectativa de la juez, que depende de los patrones de similitud perceptiva, de que otras jueces considerarán correcto su actuar, como ha ocurrido en ocasiones pasadas. Como se aclarará a continuación, la normalidad de un caso individual está asociada al reconocimiento intuitivo de una situación de familiaridad y certeza por parte de la juez, y el carácter sorpresivo de un caso está asociada a una situación de no-familiaridad e incerteza (véase Moreno Cruz, 2017: 151-213).

El Sistema 1, sobre la base de la experiencia individual de la juez, construye una representación de la normalidad por medio de la asociación intuitiva que resulta de patrones de similitud entre circunstancias, acciones y resultados que ocurren regularmente en ocasiones diferentes. Los patrones de similitud, al ser confirmados una ocasión tras otra, conforman asociaciones que determinan la creencia de la juez acerca de cómo es el presente y también su expectativa acerca de cómo será el futuro: una representación del mundo personal de la juez "*which represents what is normal in it*" (Kahneman, 2011: 71). Una situación es familiar, si el caso individual bajo examen es normal, es decir, si el caso evoca en la mente de la juez un estímulo similar a los estímulos evocados por los casos que ella ha conocido en ocasiones diferentes. Y, por el contrario, una situación no es familiar, si el caso es sorpresivo, es decir, si los estímulos evocados en su mente son diferentes a los estímulos evocados en ocasiones pasadas (Kahneman y Miller, 1986), es decir, un estímulo que la sorprende:

> *System 2 is activated when an event is detected that violates the model of the world that System 1 maintains* […] *Surprise then activates and orients your*

9 Sobre patrones de similitud perceptiva y la expectativa de que acaecimientos similares tengan consecuencias similares, véase Quine 1977: 176-196.

> *attention: you will stare, and you will search your memory for a story that makes sense of the surprising event* (KAHNEMAN, 2011: 24).

En una situación familiar, dado que el caso bajo examen en el presente no presenta propiedades diferentes en comparación con las propiedades predicables de casos similares anteriores, el caso individual evoca estímulos similares a los estímulos evocados por otros casos en ocasiones pasadas. En consecuencia, debido a la inducción intuitiva, la juez cree que, si profiere el acto habitual, sus expectativas serán satisfechas como ha ocurrido en ocasiones pasadas, en las que no se han producido resultados inesperados (MORENO CRUZ, 2018). En esta situación, la juez actúa bajo certeza. Ella no siente duda alguna acerca de cómo debe actuar, es decir, cuál acto interpretativo debe proferir para obtener con certeza el resultado representado en su intención, a saber, dictar un veredicto correcto que se deduce de la aplicación de la norma jurídica general expresada con el acto interpretativo habitual al caso individual. Un resultado que presupone que el acto proferido produzca la consecuencia –un cambio en el mundo no-intencional, pero que es objeto de una expectativa inconsciente pero accesible conscientemente– de que otras jueces considerarán correcto su actuar. La similitud perceptiva de los estímulos producidos por casos individuales, en ocasiones diferentes, influencia la intención de la juez de proferir inmediatamente un acto interpretativo habitual, puesto que ella infiere, inconscientemente, que tal acto no será cuestionado por otras jueces en el futuro, como ha ocurrido en ocasiones pasadas (MORENO CRUZ, 2014; MORENO CRUZ, 2018).

De modo que, en las situaciones familiares, de estabilidad de las creencias y expectativas conforme a un patrón de similitud y de representación de la normalidad, es decir, en relación con los casos normales, la interpretación es un mecanismo intuitivo que tiene por producto la formación de una intención de proferir inmediatamente un acto de habla interpretativo habitual (Sistema 1). Es decir, el acto interpretativo con el que ella hace uso de un enunciado interpretativo *prima facie* como la interpretación correcta de una disposición: por ejemplo, el significado textual atribuido conforme a las reglas (semánticas y sintácticas) del idioma en que tal disposición está formulada. Una interpretación que usualmente no

es argumentada, o explicitada, por cuanto que las más de las veces no es cuestionada (Moreno Cruz, 2018).

En una situación no familiar, dado que el caso bajo examen exhibe propiedades inesperadas, diferentes, anormales en comparación con las propiedades predicables de casos que sí son entre ellos similares, el caso individual evoca en la mente de la juez un estímulo que no es similar a los estímulos que otros casos individuales han evocado en su mente en ocasiones pasadas, es decir, un estímulo que la sorprende. La juez, ante casos sorpresivos, tiene razones para dejar de creer y esperar que actuar del modo habitual tenga resultados y consecuencias similares a las producidas en el pasado: es decir, ella tiene dudas acerca de que el veredicto que se deduce de la aplicación de la norma expresada con el acto habitual, al caso individual, sea un acto correcto. Y, en consecuencia, ella deja de tener la expectativa de que otras jueces consideren correcto que ella actúe del modo habitual. De modo que la juez identifica una situación de incertidumbre acerca de los resultados y consecuencias que ella puede producir con su actuar y, por lo tanto, ella se encuentra en un estado de indecisión acerca de cuál acto de habla interpretativo proferir: el habitual u otro acto alternativo opcional (ibíd.).

Luego, en las situaciones no-familiares, o de rompimiento de la estabilidad de las creencias y expectativas y del patrón de similitud entre los estímulos evocados por casos individuales en ocasiones diferentes, es decir, frente a casos sorpresivos, la interpretación es un razonamiento o una deliberación que tiene por producto la formación de una intención anticipada, es decir, un acto de toma de decisión de proferir en el futuro un acto interpretativo habitual o nuevo (Sistema 2). En estas situaciones, la juez tiene razones para dejar de creer que la interpretación *prima facie* es correcta, y, por lo tanto, para dejar de tener la expectativa de que la interpretación *prima facie* sea una interpretación que las jueces del futuro considerarán correcta. La juez, después de deliberar acerca de las opciones interpretativas, decide proferir un acto habitual –*i.e.*, usar el enunciado interpretativo *prima facie* como la norma jurídica general aplicable al caso–, o bien, ella decide proferir un acto interpretativo nuevo, *i.e.*, usar un enunciado interpretativo diferente al enunciado *prima facie* como la norma jurídica general aplicable al caso sorpresivo. Una interpretación que es generalmente argumentada por las jueces en sus sentencias (Moreno Cruz, 2018).

Hasta aquí se ha considerado que los juicios de normalidad, que son dependientes de la similitud de los estímulos evocados en la mente de la juez por casos individuales en ocasiones diferentes, son producidos por una representación de la normalidad que es construida sobre la base de consideraciones descriptivas o estadísticas, como la frecuencia y la familiaridad. En otras palabras, un caso es normal si los estímulos evocados encajan con la representación personal que la juez construye sobre la base de las asociaciones y patrones de similitud realizados por el Sistema 1, que representan lo que es normal en el mundo para la juez. Y será un caso sorpresivo si éste evoca en la mente de la juez estímulos que no encajan con dicha representación. El caso será normal si evoca estímulos que confirman las expectativas de la juez, y anormal o sorpresivo si no las confirma (KAHNEMAN y MILLER, 1986: 137). En otras palabras, un caso es normal si evoca un estímulo que comparte algunas propiedades inmutables de los estímulos evocados por casos conocidos por ella en el pasado. Y, por otra parte, cualquier propiedad mutable del estímulo evocado por el caso será percibido por la juez como anormal o sorpresivo (KAHNEMAN y MILLER, 1986: 136–153).

Sin embargo, como se expondrá a continuación, una hipótesis de la psicología cognitiva experimental es que la representación de la normalidad de casos o acaecimientos que realizamos los seres humanos está influenciada tanto por consideraciones descriptivas como por consideraciones prescriptivas.

1.2.1. REPRESENTACIÓN MIXTA DE LA NORMALIDAD

La representación de la normalidad que realizamos los seres humanos –sobre la base del Sistema 1– se suele explicar a partir de dos tipos de normas que usamos en la cognición ordinaria de ciertos casos como normales o sorpresivos: normas estadísticas y normas prescriptivas. Las primeras hacen referencia a casos que son frecuentes y no frecuentes. Las segundas requieren de un juicio de valor del caso bajo examen, más allá de su frecuencia o no frecuencia. De este modo, se ha planteado que, en relación con los juicios y usos de la palabra "normalidad", las personas hacen representaciones separadas de cada una de estas normas para proferir su

juicio. Y lo sorpresivo o lo anormal es identificado a partir de la violación de la norma estadística o de la norma prescriptiva.

Sin embargo, estudios de ciencia cognitiva recientes muestran que este no es el modo en que la mente de las personas opera realmente al hacer juicios de normalidad. En realidad, la representación de qué es lo normal está influenciado por ambos tipos de normas. Dicho en otras palabras, los juicios acerca de qué es normal –de acuerdo con el uso ordinario de esta expresión– están influenciados por juicios descriptivos acerca del "promedio" y también por juicios prescriptivos acerca de lo que creemos que sería lo "ideal". Estos experimentos muestran que la representación de lo normal está en un punto intermedio entre lo que se considera que es lo normal (promedio) a partir de normas estadísticas y lo que se considera que es lo normal (ideal) a partir de normas prescriptivas. De modo que, las personas integran información descriptiva y prescriptiva en una representación de la normalidad en la que los juicios sobre el promedio y el ideal de lo normal resultan indiferenciados (Bear y Knobe, 2017: 25-37)[10].

Y esta hipótesis, a saber, que la representación de la normalidad integra consideraciones descriptivas y prescriptivas, explica y predice cuáles son las intuiciones de las personas en relación no solo con el uso de la palabra "normal", sino también en relación con dos tipos de juicios relacionados con nuestros usos de dicha palabra. Por un lado, la identificación intuitiva del umbral –el estándar– de predicados de gradación de una clase de individuos, es decir, el punto de la escala de gradación a partir del cual se puede predicar de cierto caso individual la propiedad caliente, alto, bajo,

---

10 En uno de los experimentos se pide a un grupo de personas que juzguen cuántas horas al día es normal ver televisión. Por un lado, se les pregunta ¿cuántas creen que son las horas de televisión que las personas ven en promedio al día? La respuesta es un número de cantidad de horas que es un poco más alto que el promedio basado en estadísticas. Por otro lado, se les pregunta ¿cuántas horas de televisión sería ideal ver al día? La respuesta es un número de horas un poco más bajo que el promedio estadístico de horas de televisión vistas. Estas respuestas muestran que para las personas la cantidad promedio de horas vistas de televisión es anormalmente alto, en relación con las horas que deberían idealmente ser vistas. Y, cuando se les preguntó adicionalmente ¿cuál es la cantidad normal de horas vistas de televisión al día? La respuesta resulta ser un número intermedio entre el promedio estadístico y el ideal prescriptivo de horas vistas.

poco etc.[11]. Por otro lado, el carácter prototípico o paradigmático de ejemplificaciones de una categoría: es decir, la representación de categorías que consiste en la identificación intuitiva de algunas ejemplificaciones como *más* prototípicas que otras ejemplificaciones de la misma categoría. Al respecto, una ejemplificación es considerada una ejemplificación paradigmática, prototípica, de cierta categoría, cuando es juzgada como una instanciación más cercana al promedio de lo normal (la frecuencia, familiaridad o regularidad) y también como más cercana al ideal de lo normal (cumplir con cierto objetivo, propósito, o finalidad, asignado a la categoría)[12].

Y bien, esta hipótesis de representación mixta de la normalidad aplicada a los juicios intuitivos de las jueces en relación con la (a)normalidad de los casos individuales, implicaría reconocer la posibilidad de que los estímulos evocados en la mente de la juez están condicionados tanto por la información acerca de la frecuencia de los usos lingüísticos de las palabras de clase o predicados fácticos de los antecedentes de las disposiciones jurídicas, como por ideales de normalidad asignados, intuitivamente, a la categoría o clase de casos individuales relevante.

Retomando el famoso ejemplo en la discusión iusfilosófica: el caso individual "la entrada del carro x al parque" –donde x está por "la entrada de tanque de guerra para hacer un monumento en homenaje a los soldados caídos"–, en relación con la atribución de un significado a la disposición "la entrada de vehículos al parque está prohibida". Se trataría de un caso normal si encaja con la representación mixta de la normalidad de la juez: es decir, una ejemplificación paradigmática de "entrada de vehículo", sobre la base de una norma descriptiva acerca del uso frecuente, textual y

---

11 Bear y Knobe, 2017: si se les pregunta a las personas qué número de horas es "una gran cantidad de horas de televisión", sus respuestas confirman la hipótesis de que la identificación del "estándar" no resulta de la comparación del promedio estadístico de horas vistas con las horas que ellas ven televisión, sino de la comparación con la "cantidad normal" de horas que idealmente deberían ser vistas.

12 Bear y Knobe, 2017: la identificación intuitiva de ejemplificaciones prototípicas de una categoría depende de factores estadísticos –por ejemplo, la tendencia, la frecuencia y la familiaridad–, y también de factores prescriptivos como el de cumplir con un objetivo que es asignado a la categoría, por ejemplo, en relación con la categoría "prendas de vestir" la "necesidad de usarlas".

ordinario de la palabra "vehículo", y, también, de acuerdo con una norma prescriptiva acerca del propósito de la prohibición de la disposición: por ejemplo, los vehículos ponen en peligro la integridad física de los transeúntes en el parque, por tanto, "entrar en el carro x al parque no es entrar en un vehículo al parque" (HART, 1983)[13].

En síntesis, los juicios intuitivos acerca de la (a)normalidad de un caso individual presuponen la identificación intuitiva de una situación (no) familiar: el caso es (a)normal a la luz de una representación de la juez de la normalidad, que está influenciada por consideraciones descriptivas y prescriptivas.

A continuación, me serviré del aparato de análisis teórico-psicologista de la interpretación judicial anteriormente expuesto, con el propósito de cuestionar la tesis difundida de que las jueces no ejercen discrecionalidad en la identificación de cuál norma jurídica general aplicar a casos individuales de cierto tipo, a saber, los casos fáciles.

2. ANÁLISIS PROSPECTIVO Y RETROSPECTIVO

La interpretación judicial puede ser analizada psicológicamente en dos direcciones: prospectiva y retrospectivamente.

En dirección prospectiva, el análisis versa sobre la formación en curso de la intención de proferir un acto interpretativo. En dirección retrospectiva, en cambio, el objeto de análisis es la actividad de la juez que se activa ante la pregunta (actual o hipotética): ¿por qué proferiste ø (y no otro acto alternativo)? (MORENO CRUZ, 2017: 192-198).

Hasta este punto, hemos analizado a la interpretación judicial prospectivamente, es decir, un análisis de qué es lo que ocurre en la mente de la juez para que ella forme la intención de proferir un acto interpretativo

---

13 Hart considera que lo que hace claro a un caso es un núcleo de significado claro del texto. Y también que el propósito y el lenguaje contribuyen mutuamente en determinar el único significado de la disposición en relación con los casos paradigmáticos. En otras palabras, el propósito claro de una disposición es un factor de determinación en los casos en que el derecho se aplica claramente. El propósito de la disposición, es decir, la razón de por qué ella fue proferida, entra a determinar lo que ella significa claramente.

en relación con un caso individual. En este sentido, hemos distinguido, por un lado, una actividad intuitiva que es activada frente a los casos normales, y que tiene por producto la intención de proferir inmediatamente el acto elegido intuitivamente. Y, por otro lado, un razonamiento que se activa frente a los casos sorpresivos, y que tiene por producto la intención de proferir el acto interpretativo, habitual o nuevo, que ha sido elegido conscientemente por ella: el acto objeto de su decisión.

En esta dirección prospectiva de análisis, también se señaló que en la actividad intuitiva la juez no advierte prospectivamente un problema de decisión interpretativa. Y que, en el razonamiento de toma de decisión, por el contrario, ella sí lo advierte: ella es consciente de que hay un problema interpretativo, o, en otras palabras, ella duda acerca de cuál acto proferir entre dos o más actos interpretativos opcionales.

Sin embargo, la interpretación judicial también puede ser analizada retrospectivamente, es decir, en relación con la pregunta ¿por qué tú, juez, proferiste ø? Una dirección de análisis que es funcional para dar cuenta de la naturaleza argumentativa de la práctica de interpretar disposiciones jurídicas por parte de las jueces.

2.1. ¿POR QUÉ PROFERISTE ø?

El análisis retrospectivo, muestra el aspecto argumentativo de la interpretación judicial; en otras palabras, la actividad que la juez activa para ofrecer razones que justifiquen por qué ella profirió cierto acto interpretativo (y no otro acto posible).

En relación con el razonamiento interpretativo, dado el caso en que se interpele a la juez requiriéndole una justificación acerca de por qué ella tomó la decisión de proferir un acto interpretativo y no otro acto opcional, si, como hemos dicho, la juez advirtió prospectivamente que estaba tomando una decisión interpretativa, es decir, si ella advirtió que está pesando razones a favor y en contra de proferir uno u otro acto de habla interpretativo, entonces, ella bien podría recrear el razonamiento de toma decisión en el que ella se embarcó para disipar la duda acerca de cuál acto interpretativo proferir. Por lo tanto, ella bien podría apelar a las razones que influenciaron su decisión como premisas de su justificación. En otras palabras, ella puede recrear retrospectivamente su razonamiento y ofrecer

las razones que informaron su decisión de proferir cierto acto[14]: el acto interpretativo habitual o bien un acto nuevo.

Ahora bien, en relación con la actividad intuitiva, si, como hemos dicho, la juez no advierte prospectivamente algún problema de decisión al momento de elegir intuitivamente el acto habitual, la pregunta relevante que surge es si, no obstante, ella puede advertir el problema de decisión retrospectivamente, y, por lo tanto, puede ofrecer razones para justificar por qué actuó de ese modo y no de otro modo posible.

Para un modo de ver difundido, propuesto por H. L. A. Hart, la respuesta a esta cuestión parece ser negativa.

En su propuesta, hay casos que son fáciles, paradigmáticos, respecto de los cuales la atribución de significados a las disposiciones y la aplicación del derecho ocurre mecánica o automáticamente (es decir, en los términos psicologistas que hemos venido usando aquí, son mecanismos que ocurren intuitivamente)[15]. El significado de la disposición jurídica relevante es uno solo; no hay lugar a dudas acerca de cuál es la norma jurídica general aplicable al caso individual, por lo tanto, hay una única solución correcta de la disputa o litigio conocido por la juez; su veredicto está determinado por el significado claro de la disposición, es decir, por la única regla aplicable al caso.

En estas situaciones, según Hart, la identificación de la norma jurídica aplicable al caso por parte de la juez no consiste, en ninguna dirección de análisis, en un ejercicio de la discrecionalidad.

Para Hart, la discrecionalidad consiste en un razonamiento de toma decisión, o elección consciente, de proferir cierto acto y no otro acto opcional. El ejercicio de la discrecionalidad supone que la respuesta a

---

14 Sobre la reconstrucción retrospectiva del razonamiento para justificar una acción que se realizo después de un balanceo de razones a favor y en contra de hacerlo, es decir, una elección que es el producto del ejercicio de la discrecionalidad, véase HART, 2013: 660.

15 Es decir, en los términos usados aquí, son actividades intuitivas. HART, 1958: 607, 627: "*The general words we use* […] *must have some standard instance in which no doubts are felt about its application*". La atribución de significados y la identificación de la regla jurídica aplicable a cierto tipo de casos consiste en un acto de reconocimiento "*(in which there is little that is deliberate or even voluntary)*".

la pregunta "¿por qué elegiste proferir ø?" esté fundamentada en algún principio, objetivo, *ratio*, propósito, que compita con la opción de proferir otro acto posible que, a su vez, esté fundamentada en otro principio, objetivo o *ratio*. De modo que, el ejercicio de la discrecionalidad presupone que haya más de un principio u objetivo a tener en cuenta, y, por lo tanto, la necesidad práctica de balancear los principios en contienda, y decantar, con fundamento en uno de ellos, la decisión de proferir un acto interpretativo entre otros actos opcionales. En este sentido, el ejercicio de la discrecionalidad requiere de una argumentación que sea susceptible de aprobación o crítica racional sobre la base de los principios o estándares aducidos como fundamentos racionales de su actuar. En consecuencia, si hay ejercicio de la discrecionalidad, no hay una única respuesta correcta. En estas situaciones, en las que hay lugar al ejercicio de la discrecionalidad, el decisor puede justificar su acción reconstruyendo su razonamiento y las razones que influenciaron su decisión (HART, 2013: 652-665).

Por lo tanto, la discrecionalidad, entendida de este modo, no es aplicable a la interpretación y aplicación judicial del derecho a los casos fáciles. En relación con estos casos, hay una única norma aplicable correcta: no hay opciones interpretativas que compitan con el significado que la juez reconoce, comprende, recibe, como el único significado correcto de la disposición relevante. En pocas palabras, no hay actos interpretativos opcionales que ella pueda elegir a cambio de proferir el acto que ella elige automáticamente (ibíd., 658)[16]. Luego, la aplicación a estos casos de la única norma jurídica general expresada por la disposición –mediante un acto de habla interpretativo proferido automáticamente– es incuestionable[17].

---

16 En relación con los casos fáciles, la interpretación y aplicación del derecho no es objeto de una decisión, sino de la única elección correcta que es hecha "*to give effect to clear methods of reaching clear aims or to conform to rules whose application to the particular case is obvious*".

17 HART, 1961: 126: "*The plain case, where the general terms seem to need no interpretation and where the recognition of instances seems unproblematic or 'automatic', are only the familiar ones, constantly recurring in similar contexts, where there is general agreement in judgments as to the applicability of the classifying terms. General terms would be useless to us a medium of communication unless there were such familiar, generally unchallenged cases*".

De ello se sigue que, de acuerdo con la propuesta teórica de Hart, la interpretación y aplicación judicial del derecho en relación con los casos fáciles no es caracterizable como una actividad discrecional, ni en dirección prospectiva ni tampoco en dirección retrospectiva de análisis. Por la simple razón de que hay una respuesta correcta[18].

Ahora bien, a la luz de la perspectiva de análisis psicologista en dirección prospectiva, que haya casos individuales –llámense normales o fáciles– respecto de los que la juez no experimenta conscientemente el ejercicio de la discrecionalidad, para formar la intención de proferir automáticamente un acto interpretativo, es una creencia o afirmación teórica plausible. Por el contrario, a la luz de un análisis psicologista retrospectivo, resulta cuestionable negar el carácter discrecional de la intención que tuvo la juez de proferir ø en relación con los casos fáciles (o normales).

2.1.1. Juicios de (a)normalidad sobre actos interpretativos proferidos

Desde la perspectiva de análisis psicologista de la interpretación judicial que se ha ido proponiendo en este capítulo, la negación de que la juez pueda advertir retrospectivamente la decisión de haber proferido un acto interpretativo en relación con los casos fáciles –en otras palabras, negar el carácter discrecional de la actividad de la juez de haber proferido ø y no otro acto interpretativo opcional– conlleva implicaciones difícilmente sostenibles para la descripción y explicación de la interpretación judicial propuesta por Hart[19].

---

18 En las situaciones en las que hay una única respuesta correcta no hay lugar para hablar del ejercicio de la discrecionalidad. Se puede afirmar que se actuó correctamente, sin que tal acción haya sido elegida como producto del ejercicio de la discrecionalidad, véase Hart, 2013: 658.

19 En otras ocasiones, me he ocupado de alguna de estas implicaciones insostenibles (véase Moreno Cruz, 2017; Moreno Cruz, 2018). Aquí me ocupo tan solo de enfatizar algunas de ellas, con el propósito de caracterizar retrospectivamente a la interpretación judicial en relación con los casos individuales fáciles, como una actividad discrecional, o como el producto del ejercicio discrecional por parte de la juez.

En la propuesta de Hart, el acto interpretativo ø es incuestionable por cuanto expresa la única norma jurídica aplicable al caso, aquella que corresponde al significado textual y/o al propósito claro de la disposición (HART, 1983)[20]. Sin embargo, que el acto ø sea cuestionable o incuestionable depende de los estímulos evocados por el caso en la mente de quien lo examina –a saber, el caso de que "la juez haya proferido ø"–, y no de una zona clara del único significado de una disposición jurídica que determina de una vez por todas el carácter incuestionable del acto proferido por la juez.

Si la pregunta "¿por qué proferiste ø?" es formulada, entonces el acto ø es de hecho cuestionable. Y esto bien lo podría aceptar Hart en los siguientes términos: dado el caso en que se inquiriera a la juez, ¿por qué proferiste el acto interpretativo ø? –una pregunta que para Hart es insólita, puesto que nadie en su sano juicio cuestionaría el acto ø– la juez podría cumplir con tal requerimiento satisfactoriamente respondiendo: "porqué ø es el único acto que yo podía proferir en relación con el caso individual bajo examen"; "ø es la única cosa sensible que yo podía hacer para expresar la norma jurídica aplicable al caso"; o bien, ella podría responder simplemente con la contra-pregunta, ¿por qué no?[21].

Sin embargo, de hecho, estas respuestas son insatisfactorias en la práctica argumentativa de las jueces de pedir y ofrecer razones en contextos jurídicos. Veamos.

La pregunta "¿por qué proferiste ø?" es una forma de expresar un estado de sorpresa. Esta pregunta, que pide las razones de la juez para haber proferido ø, también presupone un juicio de (a)normalidad del acto ø que la juez profirió en relación con cierto caso individual, por parte de quien la formula: es decir, presupone la creencia de que alguna norma ha sido violada y que haber proferido ø es anormal, o, dicho de otro modo, que no haber proferido ø sería lo normal (ibíd., 148). En otras palabras, el acto ø proferido por la juez es un caso individual anormal o sorpresivo, a la luz de la representación intuitiva de la normalidad de quien inquiere. Luego, quien pone en duda la normalidad del acto ø, cree que el caso individual

---

20 Véase *supra* nota 13.

21 Sobre el uso de esta contra-pregunta para responder a ¿por qué…?, véase KAHNEMAN y MILLER, 1986: 148-149.

examinado por la juez requería que ella profiriera un acto interpretativo diferente a ø: el caso individual exhibe propiedades, atributos, relevantes que requerían que la juez lo hubiese tratado de un modo diferente al obtenido con la aplicación de la norma jurídica general expresada con el acto ø. Luego, la pregunta "¿por qué proferiste ø?" presupone que ø es sorpresivo y que haber proferido un acto diferente habría sido lo normal.

El acto interpretativo ø será anormal o sorpresivo, si la juez tenía la opción de proferir un acto diferente[22]. Y, por el contrario, el acto ø será normal si la juez no tenía otros cursos de acción alternativos, de tal manera que ø es el único acto que ella podía proferir, y, por tanto, que no proferir ø sería anormal.

En este orden de ideas, la respuesta "¿por qué no?" es satisfactoria si, y solo si, con ella la juez logra eliminar el estado de sorpresa de quién inquiere a la juez acerca de cuáles son las razones que ella tuvo para haber proferido ø y no otro acto alternativo. Esta respuesta tendría que ser suficiente para mostrar que proferir ø es normal por cuanto no hay otro acto opcional que ella hubiese podido proferir a cambio de haber proferido ø (ibíd., 139): "el acto ø es el único acto que yo, juez, podía proferir para expresar la única norma jurídica general aplicable al caso". Si esta respuesta es aceptada como satisfactoria, entonces no hay nada más qué decir para justificar por qué ø (ibíd., 148).

Sin embargo, que tal respuesta sea aceptada como razón suficiente para justificar que la juez haya proferido ø, porque era el único acto que ella podía proferir, no parece ser plausible. La pregunta "¿por qué proferiste ø?" presupone que quien la formula cree que la juez tenía la opción de proferir un acto diferente a ø. Luego, la respuesta "ø era el único acto que yo podía proferir" resultaría insatisfactoria. Para responder satisfactoriamente, la juez tiene que ofrecer las razones para justificar por qué profirió ø y, además, las razones de por qué no profirió otro acto interpretativo diferente, es decir, las razones por las cuales ella considera que ø es una mejor opción que no proferir ø y actuar de un modo diferente. Y una justificación de este tipo por parte de la juez, que es la requerida

---

22 Ibíd., 136: "*Surprise is produced in such cases by the contrast between a stimulus and a counterfactual alternative that is constructed, not retrieved*".

para responder satisfactoriamente a la pregunta "¿por qué proferiste ø?", implicaría el reconocimiento retrospectivo de un problema de decisión, es decir, del ejercicio de su discrecionalidad de haber elegido proferir ø y no otro acto diferente.

En consecuencia, que ø sea cuestionable implica que la juez tenga la capacidad de reconocer el ejercicio de su discrecionalidad retrospectivamente. Esto a pesar de que, en dirección prospectiva de análisis, se trate de un acto que la juez eligió proferir sin que ella haya ejercido su discrecionalidad o sin que la intención de proferirlo haya requerido una deliberación y elección consciente.

En conclusión, en una dirección de análisis retrospectivo, tanto los actos de habla interpretativos habituales que son objeto de la intención-presente producida por una interpretación-intuitiva –es decir, proferencias interpretativas que no involucran el ejercicio de la discrecionalidad en dirección prospectiva– como los actos habituales y nuevos que son objeto de una intención anticipada producida por un razonamiento de toma de decisión interpretativa –proferencias interpretativas que sí involucran el ejercicio de la discrecionalidad en dirección prospectiva–, son actos interpretativos que la juez eligió proferir en ejercicio de su discrecionalidad[23].

CONSIDERACIONES FINALES

En este capítulo sostuve que la distinción entre casos normales y sorpresivos es dependiente de los estímulos que los hechos del caso producen en la mente de la juez. Es decir, que un caso sea normal o anormal es dependiente del juicio intuitivo producido en la mente de la juez por los hechos del caso. De este modo, el Sistema 1, sobre la base del cual se formulan los juicios intuitivos de (a)normalidad, es el punto de partida de toda actividad cognitiva de elección práctica: la formación de la intención intuitiva o deliberada de proferir cierto acto interpretativo se da a partir

---

23 Sobre razones para actuar ofrecidas en un ejercicio retrospectivo por parte del agente, véase BLACKBURN, 1998: 125: "*it is often only in retrospect, when a course of action has settled down, that we can hit upon a story about what was influencing us, according to which our eventual decision made good sense*".

de un juicio intuitivo inicial acerca del carácter normal o sorpresivo de los casos individuales. Sin embargo, también se sostuvo que los juicios intuitivos producidos por los mecanismos inconscientes, es decir, por el Sistema 1, no siempre son el punto de llegada, es decir, el producto final del proceso de elección práctica. Hay elecciones prácticas que no son el producto de una actividad cognitiva intuitiva, sino de un razonamiento interpretativo. Frente a los casos individuales sorpresivos, el Sistema 2 se activa para considerar las razones a favor y en contra de proferir el acto de habla que es objeto de la elección práctica intuitiva.

Al respecto, vale decir algo más. Si bien la intuición es el factor que determina la formación de la intención de proferir inmediatamente los actos interpretativos habituales en relación con los casos normales, la intuición también es activa en el proceso de razonamiento que tiene por producto un acto de decisión de proferir ya sea un acto habitual o bien un acto nuevo en relación con los casos sorpresivos. La intuición, entendida como un juicio realizado sobre la base de la experiencia almacenada en la memoria de la juez, es un factor que influencia tanto la intención presente de actuar inmediatamente del modo habitual, como la intención anticipada o decisión de actuar del modo habitual o de un modo nuevo. En otras palabras, la intuición es activa tanto en situaciones familiares como en aquellas no familiares.

Frente a los casos sorpresivos, la intuición juega un rol activo en la solución del problema de decisión que la juez advierte conscientemente (véase Berger, 2013: 6). Un rol que en este capítulo no fue explorado. Al respecto, me limitaré a decir tan solo lo siguiente.

En un razonamiento realizado por expertos, profesionales, como lo son las jueces, la intuición juega un papel relevante en la identificación de los actos interpretativos opcionales del problema de decisión. La intuición es una habilidad (y no simplemente una capacidad) de los expertos para identificar cuáles son las proferencias interpretativas opcionales, en aras de solucionar la disputa, es decir, para dictar un veredicto correcto. Una vez se ha reconocido intuitivamente las opciones de actos interpretativos, estos son examinados deliberativamente por medio de la imaginación y evaluación de los resultados que cada uno de ellos puede producir en el futuro, y la simulación mental de cómo podría llevar a cabo el acto elegido,

es decir, el acto que parece funcionar suficientemente bien para obtener el resultado representado en su intención de proferirlo[24].

En este capítulo también sostuve que no hay actos interpretativos incuestionables. Que el acto proferido por la juez sea normal no depende de una zona clara del significado de las disposiciones, sino del hecho que el acto proferido por la juez no evoque estímulos sorpresivos en la mente de quien conoce y examina (por ejemplo, otra juez) el hecho de que la juez haya proferido ø. En el caso en que este hecho sí evoque estímulos sorpresivos en quien lo examina, la juez tendrá que contar con la habilidad de desvirtuar el estado de sorpresa en quien cuestiona su acto. La sorpresa requiere que la juez de razones a favor de haber proferido cierto acto y razones en contra de elegir otra opción interpretativa. Lo que presupone la capacidad de la juez de advertir retrospectivamente un razonamiento de toma de decisión interpretativa. Esto pese a ser un proceso psicológico inadvertible prospectivamente, es decir, no obstante que tal razonamiento no haya informado activamente la intención formada irreflexivamente de proferir, en situaciones de familiaridad, y en sintonía con los usos frecuentes del lenguaje, el acto interpretativo (habitual) ø.

REFERENCIAS

BEAR, A. y KNOBE, J., 2017: "Normality: part descriptive, part prescriptive", *Cognition*, 167, 25-37.

BERGER, L., 2013: "A revised view of the judicial hunch", *Legal Communication and Rhetoric: JALWD*, 10, 1-40.

BLACKBURN, S., 1998: *Ruling Passions. A Theory of Practical Reasoning.* Oxford: Clarendon Press.

BRATMAN, M. E., 1987: *Intention, Plans, and Practical Reason.* Stanford: CSLI.

HAMPSHIRE, S. y HART H. L. A. (1958): "Decision, intention and certainty", *Mind*, New Series, vol. 67, n.º 265, 1-12.

---

24 Ibíd., 3, 24.

HART, H. L. A., 1958: "Positivism and the separation of law and morals", *Harvard Law Review*, vol. 71, n.º 4, 593-629.

HART, H. L. A., 1961: *The Concept of Law*. 3.ª ed. Oxford: Oxford University Press.

HART, H. L. A., 1983: *Essays in Jurisprudence and Philosophy*. Oxford: Claredon Press.

HART, H. L. A., 2013: "Discretion", *Harvard Law Review*, 127, 652-665.

KAHNEMAN, D. y MILLER, D., 1986: "Norm theory: comparing reality to its alternatives", *Psychological Review*, 93, 2, 136-153.

KAHNEMAN, D., 2011: *Thinking, Fast and Slow*. London: Penguin Books.

MORENO CRUZ, D., 2014: "Juego de toma de decisión judicial interpretativa correcta", *Analisi e Diritto*, 237-251.

MORENO CRUZ, D., 2014a: "Three realistic strategies for explaining and predicting judicial decisions", en MATHIS, K. (ed.), *Law and Economics in Europe. Foundations and Applications*. Springer, 93-124.

MORENO CRUZ, D., 2017: "Por el camino equivocado. Carrió sobre razonamiento judicial interpretativo y casos típicos", en ROLDÁN, S. (ed.), *Homenaje a Genaro R. Carrió*. Bogotá: Universidad Externado de Colombia, 151-213.

MORENO CRUZ, D., 2018: "Interpretación en abstracto. Un fragmento del *background* de la *forma mentis* de las jueces", en COMANDUCCI, P., CHIASSONI, P. y RATTI, G. B. (eds.), *L'arte della distinzione. Scritti per Riccardo Guastini*. Vol. I. Barcelona: Marcial Pons, 137-159.

PETTIT, P., 2010: "Deliberation and decision", en O'CONNOR, T. y SANDIS, C. (eds.), *A Companion of the Philosophy of Action*. Malden-Oxford: Wiley-Blackwell, 252-258.

QUINE, W. V., 1977: "Facts of the matter", en SHAHAN, W. y MERRILL, K. R. (comps.), *American Philosophy from Edwards to Quine*. Norman: University of Oklahoma Press, 176-196.

QUINE, W. V., 1995: *From Stimulus to Science.* Cambridge-London: Harvard University Press.

RAZ, J., 1975: "Reasons for action, decisions and norms", *Mind*, 481-496, ahora en RAZ, J. (ed.), *Practical Reasoning.* Oxford: Oxford University Press, 1978, 128-143.

SEARLE, J., 2010: *Making the Social World: The Structure of Human Civilization.* New York: Oxford University Press.

SEARLE, J., 2011: "Wittgenstein and the background", *American Philosophical Quarterly*, 48, 2, 119-128.

SIMON, H., 1945: *Administrative Behavior. A Study of Decision-making Processes in Administrative Organizations.* 4.ª ed. New York: The Free Press.

VON WRIGHT, G. H., 1963: *Norm and Action. A Logical Enquiry.* London: Routledge and Kegan Paul.

WILLIAMS, B., 1981: "Internal and external reasons", en *Moral Luck, Philosophical Papers 1973-1980.* Cambridge: Cambridge University Press, 101-113.

IV.
La Razón y la razón. Psicología moral y argumentación judicial[**]

PAU LUQUE[*]

Si somos autores de nuestras vidas, lo somos solo retrospectivamente.
John Gray, *The Soul of a Marionette* (2015), p. 137

INTRODUCCIÓN

Este trabajo tiene un propósito general y programático y otro más concreto. El propósito general consiste en comenzar a conectar dos discusiones teóricas. La primera de ellas se da en el campo de la teoría del derecho o, mejor dicho, en el campo de la teoría de la argumentación jurídica. La segunda se da en el campo de la psicología moral experimental.

La primera discusión tiene que ver con la manera en que se entiende en teoría del derecho la conocida distinción entre "contexto de descubrimiento" y "contexto de justificación", que la teoría de la argumentación jurídica importó de la filosofía de la ciencia neopositivista, para dar cuenta de algunos aspectos de la argumentación jurídica, y más específicamente de la argumentación en las decisiones de los jueces. La segunda tiene que ver con la función que desarrolla la razón, o el razonamiento, en términos cognitivos, psicológicos y evolutivos sobre todo en contextos prácticos o normativos, es decir, morales, políticos y jurídicos. El objetivo, al conectar estas dos discusiones, es que la segunda de ellas, empíricamente informada, arroje algo de luz sobre la primera, no en el sentido de que ponga un punto final a la primera, pero sí en el sentido de que contribuya a enriquecerla.

---

* Investigador titular A en el Instituto de Investigaciones Filosóficas de la Universidad Nacional Autónoma de México – UNAM (México). Doctor en Filosofía del Derecho de la Universidad de Génova (Italia). Fue becario posdoctoral en la Universidad de Nápoles Federico II (Italia) y ha impartido clases en la Universidad de Girona (España), en el Instituto Tecnológico Autónomo de México – ITAM (México) y en la Facultad de Filosofía y Letras de la UNAM.

** Este trabajo fue presentado en la Universidad Autónoma de Madrid en diciembre de 2018 y en la Universitat de Girona en enero de 2019. Aunque las ideas aquí contenidas fueron recibidas con distinto entusiasmo, las observaciones en ambos lugares fueron igualmente útiles. Por otro lado, Federico Arena, Natalia Carrillo-Escalera y Fernando Rudy leyeron versiones anteriores de este trabajo y me hicieron comentarios que me ayudaron a mejorarlo. Vaya para todos ellos un agradecimiento.

El propósito más concreto de este trabajo consistirá en plantear una crítica, sobre la base de la evidencia recogida en ciencias cognitivas y psicología experimental y la interpretación que de ella han hecho algunos autores, a la concepción de la argumentación de lo que llamaré la Escuela alicantina de la argumentación jurídica. Si tal crítica resultara plausible, la consecuencia más razonable que de ella se seguiría sería la de reconsiderar esa concepción en términos más modestos.

Es necesario decir que es algo reduccionista –y, en este sentido, injusto– hablar de la Escuela alicantina de la argumentación tal y como yo lo haré aquí. Y lo es por dos razones. En primer lugar, porque la Escuela alicantina –o algunos de sus representantes, para ser más exactos– ha basculado entre dos concepciones de la argumentación que yo, en un alarde de originalidad, denominaré la concepción fuerte y la concepción débil de la argumentación. Daré evidencia textual de que, en efecto, ha sostenido ambas posiciones a lo largo del tiempo, pero me centraré, sobre todo, en la versión fuerte.

Y, en segundo lugar, es algo injusto y reduccionista porque los alicantinos no son los únicos que han sostenido esa concepción fuerte de la argumentación y, mucho menos aún, la versión débil. Pero si elijo hablar de la Escuela alicantina es porque me parece que han sido los alicantinos quienes más y mejor han reflexionado, desde un nivel conceptual de segundo orden, sobre el estatus filosófico de la argumentación. Es decir, son los alicantinos quienes, al menos en la cultura jurídica hispanohablante, más han contribuido a lo que, en sentido estricto, cabe llamar metateoría de la argumentación. Esta es la razón por la cual elijo a la Escuela alicantina de la argumentación como principal interlocutor en este texto.

Procederé del siguiente modo. Primero, presentaré de manera sintética la discusión en teoría de la argumentación jurídica. Tras ello, introduciré el debate en psicología experimental y psicología moral y daré razones por las cuales la tendencia actual en esas disciplinas empuja contra una de las dos posiciones en liza en teoría de la argumentación. Más tarde, esbozaré cuál debería ser la concepción del razonamiento que debería presuponer una teoría de la argumentación jurídica para acomodar mejor la evidencia empírica. Por último, expondré las conclusiones.

1. TEORÍA DE LA ARGUMENTACIÓN

Para los filósofos de la ciencia neopositivistas, al hablar de cómo había que entender las teorías científicas, cabía distinguir dos contextos distintos. Por un lado, está el contexto de descubrimiento. Por otro lado, está el contexto de justificación. Así, una cosa son los criterios de validación de una teoría científica (el contexto de justificación) y otra el proceso mediante el cual de hecho se llega a esa teoría (el contexto de descubrimiento); muchas veces se sostiene que ambos no tienen por qué coincidir, ya que una cosa es cuándo una teoría está justificada, *i.e.*, cuando cumple determinadas reglas o criterios, y otra cómo se descubre una teoría, *i.e.*, la cadena causal que lleva al descubrimiento científico. No siempre la cadena causal se acomoda al conjunto de reglas o criterios que justifica una teoría.

Algunos teóricos del derecho han querido importar esta distinción a la teoría de la argumentación judicial para trazar la siguiente analogía: una cosa es el proceso psicológico a través del cual el juez llega a la decisión (contexto de descubrimiento) y otra es el conjunto de razones invocadas por el juez en la parte motivacional de la sentencia (el contexto de justificación) (GUASTINI, 2014: 237; ALEXY, 1989: 416; HART, 1972: 270). No siempre el discurso justificativo coincide con el proceso psicológico (BANKOWSKI, MACCORMICK, SUMMERS y WRÓBLEWSKI, 1991: 16-17): es posible que el juez se guíe de hecho por sus sentimientos privados de justicia (sus preferencias, intereses, ideología, etc.[1]) para llegar a la decisión p, pero esos sentimientos privados no figurarán en la sentencia, dado que, como dice Guastini, p debe estar fundada en alguna norma jurídica positiva (GUASTINI, 2014: 238), ya sea ordinaria o constitucional.

Así, en la sentencia cuya decisión es p la norma jurídica positiva de turno figurará como premisa normativa, pero esto no quiere decir que de hecho el juez esté motivado por esa norma a la hora de decidir p; lo que

---

1 "Los escépticos jurídicos sostienen no solo que el significado no queda fijado al promulgar una ley, sino que evoluciona en el proceso de entender y aplicar el derecho. Más aún, creen que semejante proceso es extremadamente dinámico e influenciado, si no es que enteramente determinado, por factores subjetivos y emocionales" (KLATT, 2008: 22; trad. propia). En este sentido, véase KENNEDY, 1996, o SCHAUER, 2009: 4.

quiere decir es justamente que en el argumento que justifica *p* esa norma jurídica positiva es una premisa normativa sin la cual no es posible justificar *p*. En este sentido, es posible que el argumento cuya conclusión es *p* no fuera nada más que una justificación *ex post*, es decir, el juez primero toma la decisión y luego rebusca en el derecho positivo qué norma le sirve para justificar *p*.

Está lejos de ser obvio que sea apropiado trasladar la distinción entre contexto de descubrimiento y contexto de justificación, tal y como es entendida en filosofía de la ciencia, a la teoría de la adjudicación (Mazzarese, 1995). Pero me parece, en cualquier caso, que esa distinción captura un punto interesante y relevante acerca de la decisión judicial, esto es, la posible disparidad entre las razones que justifican (en sentido normativo) la decisión de un juez y las que la motivan (en sentido psicológico). En este sentido, a veces la distinción entre contexto de descubrimiento y contexto de justificación es entendida como la diferencia entre razones explicativas –las que motivan al juez– y razones justificativas –las que determinan, en sentido normativo, la corrección– de la decisión judicial (Atienza, 2014: 114; Aguiló, 1997: 73).

Dos ilustres representantes de la Escuela alicantina de la argumentación, Manuel Atienza y Josep Aguiló, han afirmado en diversos lugares que esa querencia –muchas veces abrazada por filósofos del derecho que albergan algún grado de escepticismo respecto de la determinación del derecho– por distinguir esos dos momentos es, en el mejor de los casos, exagerada[2]. Correlativamente, también la distancia entre las razones explicativas y las razones justificativas habría sido achicada. Reconstruyo a continuación de forma sintética sus ideas al respecto.

Según Atienza, "es perfectamente posible que una razón explicativa tenga también fuerza justificativa; al fin y al cabo, el buen juez es aquel

---

2 Aunque en otros lugares, singularmente Atienza, afirma la vigencia y utilidad de la distinción. No cito en qué lugares porque, en el caso de Atienza, la aceptación de la distinción parece ser más frecuente que su impugnación (al contrario de lo que ocurre, como veremos, en el caso de Aguiló), por lo que me parece que tiene más valor citar aquellos pasajes –como hago en el texto– en que se desmarca de la posición, por llamarla de alguna manera, por defecto.

que toma sus decisiones guiado únicamente por *buenas* razones" (Atienza, 2014: 115; las cursivas son mías). Creo que esta es una buena manera general e inicial de expresar la posición de la Escuela alicantina. Pero me parece que merece ser más concretada.

Yo diría que para los alicantinos la vigencia de la distinción depende, por un lado, de cuál de las tres dimensiones (formal, material y pragmática) de la concepción argumentativa estemos hablando y, por otro lado, de si "argumentación" es interpretado como resultado o como actividad (dado que padece de la ambigüedad proceso/producto).

Desde un punto de vista formal, la distinción puede ser mantenida, creo –y así es ampliamente aceptado–, sin mayor problema: se puede reconstruir el esquema formal de la fundamentación de una decisión judicial atendiendo únicamente al contexto de justificación (Atienza, 2014: 115). La concepción formal está interesada en la argumentación como resultado. No importa cómo se pase de hecho de las premisas a la conclusión, lo importante es que la decisión se pueda reconstruir de modo tal que no haya ninguna inferencia ilegítima desde el punto de vista lógico.

> Y así, por ejemplo, el hecho de que, desde el punto de vista psicológico, se alcance primero la conclusión y luego se formulen las premisas mediante un mecanismo de 'racionalización' a posteriori, no afecta para nada a la cuestión de si el paso –la inferencia– está o no justificada lógicamente (Atienza, 1999: 46).

Dicho de otro modo, se pueden satisfacer las exigencias derivadas de la concepción formal de la argumentación incluso cuando las razones explicativas y las razones justificativas no coincidan; lo que se necesita es poder reconstruir la decisión con la forma de un argumento lógicamente válido.

Desde un punto de vista material, hay –aparentemente– por lo menos dos maneras de entender una buena argumentación: 1) Una argumentación correcta es una argumentación válida y a la vez sólida, con independencia de que el decisor esté motivado por las buenas razones, es decir, con independencia de cuál haya sido el proceso psicológico por el que el juez haya concluido *p*; 2) Una argumentación correcta es una argumentación válida y a la vez sólida y en la que el decisor está motivado por las buenas razones, es decir, por las razones que convierten a las premisas en

verdaderas y a la conclusión en un medio idóneo para la persecución de fines sociales valiosos. 1) ve la argumentación como resultado, mientras que 2) la ve como una actividad. De este modo, si se ve la argumentación como actividad, una argumentación válida y sólida en la que su decisor no está motivado por las buenas razones no será una argumentación correcta. Pero si se ve la argumentación como resultado, entonces basta con que se pueda reconstruir la decisión como un argumento lógicamente válido y en el que sus premisas sean verdaderas u objetivamente correctas para que se trate de un argumento correcto desde el punto de vista material (es importante señalar que la concepción material de la argumentación presupone la formal, pero no viceversa).

La Escuela alicantina tiende a ver la argumentación como una actividad (Atienza, 1999: 47), por lo que parece que la versión preferida por ellos de la concepción material sería 2), es decir, una buena argumentación sería aquella en que no hay distinción entre contexto de descubrimiento y contexto de justificación porque la sentencia refleja que en el proceso psicológico del juez para tomar la decisión figuran las mismas razones que en la sentencia. Así, la concepción material de la argumentación entendida como actividad diría que un argumento es correcto si es válido, sólido y el argumentador está motivado por las razones correctas[3].

Desde el punto de vista pragmático, la distinción entre contexto de descubrimiento y contexto de justificación prácticamente desaparece, ya que el criterio de "evaluación de los argumentos es aquí inseparable del propio comportamiento de los participantes y/o de los destinatarios" (Atienza, 1999: 47). Una argumentación es correcta desde el punto de vista pragmático en función de que la misma sea aceptada por los involucrados (tanto el auditorio como el "autor" de la argumentación), es decir, en función de cuáles sean las actitudes de los participantes. Y dado que

---

3 Por ejemplo, no estaría motivado por las razones correctas si él decidiera *p* no porque es correcto (en el sentido de que es la conclusión de un argumento válido y sólido) sino, por ejemplo, porque sabe que decidir *p* le ayuda a mejorar su carrera judicial o porque la presión social a favor de *p* es tan grande que hace que decidir de otra manera sea un peligro para su integridad física. En estos dos supuestos no se trataría de una argumentación correcta con arreglo a la concepción material de la argumentación (como actividad).

la concepción pragmática presupone la concepción formal y también la concepción material (entendida como una actividad), se sigue que los jueces deben aceptar, si es que queremos una argumentación correcta, la decisión por las razones correctas[4].

Por su lado, Josep Aguiló también ha puesto énfasis en que el hiato entre las razones explicativas y las razones justificativas, o entre contexto de descubrimiento y contexto de justificación, es –o debería ser– mucho menor de lo que afirman los escépticos.

Según Aguiló, el ideal del Estado de derecho impide esa separación tajante: el principio de independencia judicial y el principio de imparcialidad hacen que el derecho no se pueda desentender del contexto de descubrimiento (Aguiló, 1997: 77-78). En ese sentido, estos principios hacen que las razones explicativas tengan que coincidir con la justificativas (Atienza, 2014: 116; Aguiló, 2009: 29), y con ello se reafirma que una buena argumentación, en sentido material, debe ser entendida no solo como resultado, sino como actividad.

Veamos con un poco más de detenimiento el argumento. Aguiló invoca a Kant para afirmar que llevar a cabo una conducta moral no consistía, a ojos de Kant, en obrar

> … en correspondencia con el deber, es decir, que la conducta se adaptara al contenido del deber; sino en obrar por reverencia al deber, esto es, movido por el deber. No es difícil observar que ello quiere decir que, conforme con Kant, en el ámbito de la conducta moral, la razón explicativa –el motivo de la conducta– y la razón justificativa –el deber– coinciden (Aguiló, 1997: 73-74).

---

4 Aquí, de nuevo, parece en principio plausible que haya participantes que acepten esa decisión motivados por razones puramente instrumentales y no porque esa decisión sea la correcta. De suceder esto con alguna argumentación, esta no podría ser considerada correcta –aunque obtuviera aquiescencia– según la concepción pragmática tal y como ha sido delineada en el texto. Una argumentación persuasiva no es una argumentación a la cual los interlocutores se adhieren por la razón que sea, es una argumentación a la que se adhieren porque es la argumentación correcta.

A mi juicio, en muchos casos judiciales esta noción de argumentación correcta parece sobreexigente, pues significa que la argumentación será correcta solo si es aceptada por las razones correctas incluso por la parte que es derrotada en el caso, algo que parece improbable.

El principio de independencia es la manera en que el derecho exige obediencia a sus jueces; "los deberes del juez no son tanto límites a su poder, cuanto determinaciones positivas de su conducta" (ibíd.: 75). En este sentido, el juez no es autónomo: el juez debe ser independiente incluso de sus propios credos autónomamente aceptados (ibíd.: 76). El principio de independencia no solo protege las razones que se alegan en la sentencia, "sino que además exige que el juez falle por las razones que el Derecho le suministra" (ibíd.: 77).

El principio de imparcialidad, por su lado, está conectado con las instituciones procesales de la abstención y la recusación. Entre las causas que habitualmente califican para la abstención y la recusación están el parentesco, la amistad íntima, la enemistad manifiesta, etc. A todas ellas se les reconoce "una extraordinaria fuerza motivacional de la conducta" (ibíd.: 77), por lo que aquello que el juez hace cuando se abstiene es que si no se abstuviera "sus decisiones podrían ser vistas como motivadas por razones distintas a las suministradas por el Derecho" (ibíd.: 78). De todo lo anterior se sigue, siempre según Aguiló, que no es cierto que al derecho, o a la teoría de la argumentación jurídica, le interese solo el contexto de justificación: el derecho trata de controlar los motivos por los cuales los jueces actúan; en el ideal del Estado de derecho de un juez independiente e imparcial "hay algo muy parecido a la exigencia kantiana para la conducta moral, pero referido al marco institucional del Derecho: que la explicación y la justificación de la conducta coincidan" (ibíd.: 79).

Tras estas dos breves reconstrucciones, considero que se puede decir que la Escuela alicantina parece estar comprometida con la tesis de que concebir el derecho como argumentación significa someter el poder a la razón (Atienza, 1999: 38), en el sentido de que, al fusionar (o por lo menos acercar mucho) el contexto de justificación y el contexto de descubrimiento, la Escuela alicantina parece estar atribuyendo a la razón la capacidad no solo de justificar decisiones correctas, sino de motivarlas, volviéndose inmune al poder –digamos, de los intereses políticos– que pueden ejercer entidades externas a los jueces o los jueces mismos. La razón sería, en fin, el instrumento que dirigiría a los jueces a descubrir las respuestas correctas (o justificadas), o por lo menos las mejores respuestas, en sus sentencias. En este sentido, "una decisión judicial de contenido correcto (aplicación correcta de la ley) pero tomada por motivos incorrectos resulta,

en términos normativos, inaceptable, inasumible" (AGUILÓ, 2009: 32). La concepción de la argumentación de la Escuela alicantina parece presuponer la plausibilidad de la noción kantiana de razón.

Es importante hacer notar que los alicantinos no han sostenido este acercamiento tan contundente entre razones explicativas y razones justificativas en todos los lugares. En ese sentido, como anuncié anteriormente, podría decirse que existe una versión más débil de la Escuela alicantina y otra más fuerte. En la versión más débil, no habría acercamiento entre contexto de justificación y contexto de descubrimiento y, por lo tanto, parecería que la corrección material de un argumento judicial sería independiente de la actitud psicológica o motivacional del juez.

Si uno toma la versión más débil, entonces cabe la posibilidad de reinterpretar algunos pasajes ambiguos de los alicantinos apuntando no a la reducción de la distancia conceptual entre razones explicativas y razones justificativas, sino a la necesidad, en el marco del ideal del Estado de derecho, de que los jueces estén motivados por las buenas razones (las razones correctas de derecho). Quizás esto configuraría un marco más exigente del Estado de derecho, es decir, no estaríamos ante un verdadero Estado de derecho si los jueces, a pesar de tomar las decisiones correctas, no estuvieran motivados por las razones de derecho correctas. Pero esta sería una tesis normativo-institucional referida al Estado de derecho.

En la versión fuerte, en cambio, la exigencia no es, o no solo es, institucional, sino conceptual o lógica: la argumentación, desde el punto de vista material, no es correcta si el juez no está motivado por las razones de derecho correctas.

Yo tomaré aquí en cuenta la versión más fuerte de la Escuela alicantina. Se trata de la más original –por lo menos en teoría de la argumentación judicial– y ambiciosa.

2. LA CRÍTICA A LA *RAZÓN*

Me parece que no es baladí que la Escuela alicantina de la argumentación apele a Kant para dar fuerza a su concepción de la argumentación. Los alicantinos parecen fundar buena parte de su concepción argumentativa en una gran confianza en la razón, como lo hacía el propio Kant, entendida como aquella herramienta que nos permitiría descubrir verdades

normativas (ya sean éstas morales o jurídicas) que se articularían mediante la argumentación.

Me parece que es justamente la razón en este sentido la que permitiría acercar hasta prácticamente hacer colapsar la distinción entre razones explicativas y razones justificativas, ya que la razón –de impronta o aroma kantiano– permitiría *descubrir* las decisiones *correctas* (y, con ello, también el colapso entre contexto de descubrimiento y contexto de justificación).

En los últimos años ha aparecido una importante y voluminosa literatura proveniente del campo de la psicología experimental, empíricamente informada, que trata de aportar algo de luz sobre la manera en que los humanos razonamos. Es posible que buena parte de la motivación que explique el auge de esta literatura descanse en una perplejidad: si la razón es un instrumento programado para alcanzar el conocimiento y la verdad, como querría la gran familia kantiana –aunque no solo ella–, ¿qué explica los sesgos, los usos interesados de la misma, las falacias argumentativas frecuentes, etc.? Es decir, ¿qué explica que en muchas ocasiones la razón no amplíe nuestro conocimiento ni sea usada para proporcionarnos la verdad?

Y la respuesta, en general, de esa literatura es que la razón no parece estar evolutivamente conservada para alcanzar la verdad y ampliar nuestro conocimiento[5]. A la concepción kantiana de la razón, que denominaré, siguiendo a Mercier y Sperber (2017), *Razón*, se contrapone la concepción de la *razón*, para la cual la función principal de la misma no es descubrir verdades, sino encontrar justificaciones *ex post* para nuestros comportamientos y nuestras decisiones y, así, intentar persuadir a nuestros interlocutores.

Las teorías críticas con la concepción kantiana difieren entre ellas. Pero todas comparten la idea de que la evidencia empírica no parece respaldar la *Razón*[6].

---

5 Yo me basaré aquí sobre todo en algunas de las propuestas de Mercier y Sperber, 2011, 2012 y 2017; Haidt, 2012, y Haidt y Bjorklund, 2008. Todas ellas coinciden en la visión que acabo de enunciar en el texto.

6 Quiero mantener bajo control las consecuencias de este trabajo y por ello quiero advertir que no estoy afirmando que de la evidencia empírica no se sigue, en sentido de implicación lógica, la *Razón*, sino que esta última no está apoyada en evidencia empírica. Nada más –pero tampoco nada menos– que esto.

La teoría de Jonathan HAIDT (2012), que él mismo denomina intuicionista, extrae de una serie de experimentos[7] la conclusión según la cual el juicio moral está principalmente causado por un proceso cerebral automático –al que Haidt denomina "intuición"– y no por lo que comúnmente llamamos razonamiento (ibíd.: 54). La secuencia temporal fenoménica del modelo intuicionista sería entonces la siguiente:

> [Haidt] evento disparador//intuición//juicio moral//razonamiento.

Primero se da un evento en el mundo (es decir, un hecho, que puede ser natural o social). Ese evento causa una reacción automática en la persona: una intuición favorable, desfavorable o de indiferencia hacia ese evento. Sobre la base de esa intuición, la persona construye el juicio moral acerca de ese evento. Aunque Haidt no lo dice con las palabras que voy a usar, creo que podría afirmarse que, en este modelo, el juicio moral es la verbalización de la intuición. Tras esa verbalización, la persona alega las razones por las cuales sostiene ese juicio moral, es decir, argumenta a favor de ese juicio moral[8].

Lo que sugiere la evidencia y afirma Haidt, por tanto, es que la moralidad no proviene del razonamiento (HAIDT, 2012: 31). Las intuiciones son

---

7 Para saber en qué consistían esos experimentos, véanse los primeros capítulos de HAIDT, 2012, y HAIDT y BJORKLUND, 2008. Aquí me limitaré a reconstruir las conclusiones que Haidt extrae de sus experimentos.

8 Sin embargo, la moral es una empresa social y, como tal, podría ocurrir que las razones que invoque un interlocutor disparen nuevas intuiciones (es decir, podría ocurrir que el interlocutor que da sus razones fuera el evento disparador de nuestra intuición). El modelo intuicionista admite que esto podría ocurrir, pero solo mostraría que las intuiciones pueden estar generadas por interacciones sociales, no, en sentido estricto, por las *razones* invocadas por el interlocutor (HAIDT, 2012: 55). Es más, no es posible, siempre según el modelo intuicionista, "cambiar las formas de pensar de las personas refutando completamente sus argumentos" (ibíd.: 57; trad. propia).
Para dudar del modelo intuicionista tendría que ser el caso que la gente cambiara su juicio moral porque sus argumentos quedaran desacreditados, pero lo que la evidencia parece mostrar es que la gente cambia de juicio moral cuando se disparan nuevas intuiciones, no cuando el razonamiento que sostenía el juicio moral precedente es atacado.

una suerte de *gut feelings* que orientan o disparan el razonamiento (ibíd.: 30), pero la intuición no depende del éxito o fracaso del razonamiento (ibíd.: 51).

Si los juicios morales no son causados por el razonamiento moral, ¿cuál es entonces el rol del razonamiento? Según Haidt, y como ya dejé entrever un poco más arriba, el razonamiento moral es una fabricación *post hoc* (ibíd.: 30) destinada a justificar –no a descubrir– el juicio moral. En este mismo sentido, Margolis afirma que los juicios están producidos por una maquinaria no-consciente en el cerebro y que los seres humanos producimos *rationales* que creemos que dan cuenta de esos juicios; pero esos *rationales* son únicamente racionalizaciones *ex post* (Margolis, citado ibíd.: 50). La razón –dice Haidt de forma un poco provocadora– es la sirvienta de la intuición (ibíd.: 54).

De este modo, el razonamiento moral no busca descubrir ninguna verdad moral (ibíd.: 59), sino que busca justificar el juicio moral derivado de la intuición y, cuando la persona invoca las razones para ello, intenta que los demás compartan ese mismo juicio (ibíd.: 52).

El razonamiento moral es "meramente" estratégico: busca ganar amigos e influir en la gente (ibíd.: 53). En este sentido, cuando razonamos conscientemente buscamos persuadir a los demás. Por todo ello, Haidt concluye que el pensamiento moral se asemeja mucho más al de un político a la caza de votos que al de un científico que busca acceder a la verdad objetiva (ibíd.: 89)[9].

---

9 Es importante señalar que, según Haidt, la intuición (y el razonamiento moral) es un proceso cognitivo. En este sentido, la distinción importante no es –como normalmente se asume en metaética– entre estados mentales cognitivos y no-cognitivos, sino entre dos tipos diferentes de cognición: la intuición y el razonamiento (Haidt, 2012: 53). Esto, para quienes estén acostumbrados a leer metaética, podría llevar a confusión, porque podrían llegar a pensar que, en la medida en que tanto el juicio moral como el razonamiento moral son procesos cognitivos, admiten una semántica veritativo-funcional. Sin embargo, me parece que es únicamente un problema de vocabulario: para los psicólogos morales –o al menos para Haidt– "cognitivo" no está relacionado con una semántica veritativo-funcional (probablemente a Haidt ni siquiera le interese la discusión moral en el nivel semántico), para los que se dedican a la metaética, en cambio, sí. Por ello, cuando Haidt dice que la distinción importante es la que se da

Además, para Haidt esa función del razonamiento moral –como afirmé más arriba– tiene una explicación evolutiva. Es más adaptativo usar nuestros cerebros para dar con aquellas "pruebas" que dan apoyo a lo queremos creer que para buscar la verdad. Para nuestros ancestros, y para nosotros mismos, aumenta más nuestras probabilidades de supervivencia y reproducción persuadir (y con ello tener una "buena" reputación en un contexto social) que buscar la verdad (ibíd.: 83, 87). El razonamiento fue diseñado por la selección natural no como un método de acceso a la verdad moral, sino como una manera de justificar nuestros juicios morales (ibíd.: 86, 145)[10].

La teoría de Hugo Mercier y Dan Sperber (Mercier y Sperber 2011, 2012, 2017) no está centrada únicamente en el razonamiento en la dimensión moral, sino en un lugar más amplio respecto del razonamiento en general. Y aunque hay algunas diferencias notables con la teoría de Haidt –de las que hablaré más tarde–, la teoría de Mercier y Sperber coincide, a grandes rasgos, en el punto clave según el cual el razonamiento tiene dos funciones principales y evolutivamente conservadas: la de producir razones para justificarse a uno mismo y la de producir argumentos para convencer a los otros (Mercier y Sperber, 2017: 8, 124). Además, coinciden con Haidt en señalar que las razones son "racionalizaciones *a posteriori*" (ibíd.: 109; trad. propia) para nuestras intuiciones, sin embargo aquellas no juegan un papel a la hora de explicar el proceso inferencial que nuestra cerebro realiza y que es experimentado como una intuición (ibíd.: 117).

Si fuera verdad que el razonamiento fuera un instrumento destinado a mejorar nuestras creencias y a tomar decisiones correctas, entonces debería estar orientado hacia aquella información que violara nuestras

---

entre intuición y razonamiento, entendidos como dos tipos de cognición, simplemente está diciendo, creo, que la intuición es un proceso automático y el razonamiento un proceso reflexivo o consciente. Pero ninguno de los dos procesos rastrea verdad moral alguna (siempre según el intuicionista) y, por tanto, ninguno de los enunciados que corona cada uno de estos procesos es veritativo-funcional.

10 En este sentido, nacemos programados para pensar en términos morales (es decir, la moral es innata), pero el contenido concreto de nuestro pensamiento moral (qué es correcto o incorrecto moralmente) no es seleccionado naturalmente, sino socialmente: depende sobre todo del grupo social y cultural al cual pertenecemos (Haidt, 2012: 31).

expectativas acerca de cómo es el mundo. El razonamiento debería buscar contraargumentos para nuestras generalizaciones o cuestionar nuestras decisiones. Sin embargo, esto no es lo que sugiere la evidencia: no solemos buscar razones a favor y en contra de nuestras decisiones o preferencias en la vida normal; puestos en el brete, solo solemos buscar razones a favor de nuestras decisiones (ibíd.: 210, 213, 218).

Este rasgo o esta tara del razonamiento ha sido denominado a veces "sesgo de confirmación". Pero según Mercier y Sperber no tiene sentido llamarlo así, porque si la función evolutivamente conservada del razonamiento, al producir razones, es la de justificar nuestras propias acciones o convencer a los demás, entonces debería tener un "sesgo a nuestro favor", y no exactamente un "sesgo de confirmación" (ibíd.: 219). En este sentido, el hecho de que busquemos razones a favor de nuestras decisiones o elecciones, y solo a favor de ellas, no debería ser interpretado como un defecto cognitivo de nuestro razonamiento, sino como aquello en lo que consiste el razonamiento mismo y para lo cual, insisto, ha sido evolutivamente conservado (Mercier y Sperber, 2011).

Tenemos una inagotable habilidad para producir razones para justificar prácticamente cualquier cosa que creamos (Mercier y Sperber, 2017: 252); los humanos somos auténticas máquinas de racionalización (ibíd.: 253)[11].

La principal función del razonamiento, por tanto, es social (ibíd.: 176): ayuda a resolver problemas de coordinación al producir justificaciones, y problemas de comunicación produciendo argumentos (ibíd.: 183).

La conclusión de Mercier y Sperber es que la *Razón*, entendida como la capacidad de procurar mejor conocimiento y decisiones, no ha sido evolutivamente conservada. Lo que ha sido conservado es una capacidad más modesta: la capacidad para producir razones para justificarnos a nosotros mismos y para intentar persuadir a los demás, a saber, lo que ha sido evolutivamente conservado es la *razón* (ibíd.: 200–201).

---

11 Mercier y Sperber repiten la misma idea varias veces: "Pensar acerca de las razones a favor de la decisión que tengo que tomar es algo que normalmente ocurre después de haber tomado esa misma decisión" (Mercier y Sperber, 2017: 123; trad. propia); "Inferimos cuáles deben haber sido nuestras razones a partir de las conclusiones a las que hemos llegado intuitivamente" (ibíd.: 142; trad. propia).

Así, si buena parte de la concepción de la argumentación más fuerte de la Escuela alicantina descansa en la plausibilidad de la *Razón*, y si la evidencia empírica sugiere fuertemente que operamos con *razón*, y no con la *Razón*, entonces esa concepción de la argumentación quedaría por lo menos en parte desacreditada, dada la ausencia de evidencia empírica a su favor. Como consecuencia, el acercamiento entre el contexto de descubrimiento y el contexto de justificación, que parecía presuponer la plausibilidad de la *Razón*, quedaría en parte noqueado.

Si volvemos a la dos concepciones de la argumentación brevemente presentadas en la sección anterior, parece que la evidencia empírica apoya aquella que normalmente es calificada como escéptica. Esto no significa –algo que desarrollaré brevemente en la siguiente sección– que la argumentación sea una empresa puramente retórica, casi cínica. La implausibilidad de la *Razón* no convierte a la argumentación en algo desquiciado; lo que ocurre es que debe ser reconsiderada en los términos de la *razón*.

3. A FAVOR DE LA *RAZÓN*

Es importante señalar que hay algunas diferencias notables entre la teoría de Haidt y la de Mercier y Sperber. La teoría de Haidt, al igual que la de Kahneman y Tversky, es dualista, porque afirma la existencia de dos sistemas de decisión y razonamiento humanos: la intuición (o sistema S1 en el caso de Kahneman y Tversky) y el razonamiento (o sistema S2 en el caso de Kahneman y Tversky). El primero toma decisiones de forma casi automática, el segundo, en cambio, opera de manera más lenta y pausada. Para Haidt, como ya dije, el razonamiento tiene una función básicamente social y justificativa y es casi imposible que altere un juicio moral, con lo que se quiere decir que difícilmente el razonamiento modificará alguna de nuestras intuiciones.

Mercier y Sperber son, en este punto, menos radicales. Por un lado, consideran que las concepciones duales son erróneas. Ellos afirman que existe un único sistema de toma de decisiones: una serie de módulos biológicos que procesan información y hacen inferencias. Se trata de inferencias que son experimentadas como intuiciones. Pero el razonamiento no es una alternativa a la inferencia intuitiva: "razonar es hacer uso de las inferencias intuitivas acerca de las razones" (MERCIER y SPERBER, 2017: 133; trad.

propia). Algunos módulos captan regularidades y arrojan intuiciones. Otros módulos, en cambio, nos permiten "derivar razones para nuestras mismas intuiciones a través de otro proceso retrospectivo de inferencias intuitivas. Inferimos cuáles deben haber sido nuestras razones a partir de las conclusiones a las que hemos llegado de forma intuitiva" (ibíd.: 142; trad. propia). Así que el razonamiento no es una empresa independiente, porque si lo fuera significaría que una razón tendría que estar siempre basada en otra razón de orden superior, y esta última, a su vez, tendría que estar basada en otra razón de orden superior, con lo que un regreso al infinito afloraría. El reconocimiento de una razón está basado en última instancia en una inferencia intuitiva, y esta última está basada a su vez en las regularidades del mundo (ibíd.: 121-132). Así es como se interrumpe la cadena del regreso infinito. La diferencia fundamental entre intuiciones y razonamiento, en el caso de Mercier y Sperber, no es lógica, sino fenoménica: las inferencias que capturan y procesan regularidades acerca del mundo son experimentadas como intuiciones; las inferencias acerca de esas inferencias, es decir, el razonamiento, no son experimentadas como intuiciones. Pero esto no quiere decir, como pretenden los dualistas, que el razonamiento y la intuición sean, en términos lógicos (y posiblemente también ontológicos), dos sistemas independientes, simplemente la "*Razón* es un módulo de inferencia entre otros" (ibíd.: 327; trad. propia).

Pero existe otra diferencia notable, y más relevante a los efectos de la discusión sobre la argumentación, con la teoría de Haidt. Mientras este último no da al razonamiento prácticamente ningún rol en la conformación de nuestros juicios morales, que recae por completo (o casi) en nuestras intuiciones, Mercier y Sperber, a pesar de que creen que la principal función del razonamiento es social y no epistémica, son algo más optimistas a este respecto. En determinadas circunstancias, la *razón* puede hacernos cambiar de opinión (ibíd.: 307). ¿Cuáles son esas circunstancias? Mercier y Sperber llaman a su enfoque interaccionista, en contraposición al enfoque kantiano, al que llaman intelectualista (nótese que esta contraposición es, a grandes rasgos, paralela a la de *razón* vs. *Razón*).

Para el enfoque interaccionista existe una división del trabajo cognitivo: al producir argumentos, deberíamos apoyarnos en el "sesgo a mi favor"; en el momento de evaluarlos deberíamos ser objetivos y demandantes (ibíd.: 331). O, dicho de otro modo: en nuestro soliloquio interno, el razonamiento

debe estar sesgado; en la interacción social, en cambio, el razonamiento, expresado a través de la argumentación, debe mostrar debilidades en lo sostenido por los demás. Esto no significa que siempre que se da la interacción social, el intercambio de razones, se dé el cambio de opiniones. Solo significa que es más probable que ocurra en un entorno social que en un estado solipsista.

Pero tampoco vale cualquier tipo de entorno social. Ha sido ampliamente estudiado el fenómeno de la polarización de grupos, en el sentido de que la interacción en un grupo en el que todo el mundo piensa más o menos igual es más o menos como pensar en solitario (ibíd.: 244). Se requiere un escenario de interacción en que concurran diversas opiniones para que la *razón* pueda alterar algunas de ellas.

Así pues, la *razón* tiene como función principal la de producir razones que ayuden a nuestra reputación social. Sin embargo, en un entorno social en que hay pluralidad de opiniones a veces la *razón* puede contribuir a modificar aquellas. Esto no quiere decir que en última instancia los juicios morales no estén fundados sobre intuiciones, porque, como hemos visto, para Mercier y Sperber el razonamiento no es sino una forma de inferencia intuitiva acerca de otras inferencias intuitivas. Pero sí quiere decir que, a diferencia de Haidt, no habría que ser fatalmente pesimistas respecto de la capacidad del razonamiento para influir en nuestros juicios morales[12]. Aunque la función de la *razón* sea básicamente social, existe un espacio para que la *razón* influya en la conformación de juicios morales.

Por lo demás, es importante señalar dos observaciones más de la teoría de Mercier y Sperber. En primer lugar, a veces las personas no encontramos buenas razones para lo que estamos a punto de hacer y cambiamos nuestro curso de acción y nuestra decisión. En tales casos, sostienen Mercier y Sperber, las razones personales invocadas "han jugado un papel de rol causal en su decisión final. Estas razones no moldearon la decisión que fue finalmente tomada, pero cohibieron que la decisión inicial fuera llevada a cabo" (ibíd.: 125; trad. propia). Pero de esto no se sigue que el proceso causal vaya desde las razones hacia la decisión, como querrían los partidarios de la *Razón*

12 Esto quiere decir que aunque es cierto que Haidt rechaza la *Razón*, no parece aceptar, no al menos con todas sus consecuencias, la *razón*.

(y como probablemente pensamos, equivocadamente, todos los que nos consideramos racionales), sino que va desde las decisiones tentativas hacia la búsqueda de una justificación, el fracaso de esa búsqueda y la decisión revisada para tener un curso de acción o de decisión justificado de mejor manera (ibíd.: 125-126). Dicho de forma distinta, que el razonamiento no nos conduzca a la verdad no quiere decir que únicamente pueda ser usado a efectos retóricos para justificar prácticamente cualquier cosa.

En segundo lugar, conviene señalar que si la *razón* parece ser el caso, al menos según lo que sugiere la evidencia empírica, y además el acceso a nuestras razones explicativas parece estar –en sentido estricto– vetado incluso a la introspección (véase WILSON, 2002), entonces ese acercamiento radical y kantiano entre contexto de justificación y contexto de descubrimiento que pide la Escuela alicantina de la argumentación debería ser revisado o por lo menos bastante relativizado, como ya mencioné. Esto no quiere decir, sin embargo, que la desconexión entre razones explicativas y razones justificativas sea total. De hecho, las razones justificativas, según Mercier y Sperber, tienden a iluminar factores que jugaron un rol causal en nuestras mentes: "Las razones normalmente están construidas a partir de pedacitos de comprensión psicológica" (ibíd.: 126; trad. propia). Interpreto esta frase en el sentido de que no parece descabellado suponer que lo que decimos para justificarnos y para convencer a los demás está de algún modo conectado con aquello que nos motiva, aunque no tengamos acceso epistémico a esa conexión. Pero nótese que este es un acercamiento lateral y más modesto entre contexto de justificación y contexto de descubrimiento que el que sugerían Atienza y Aguiló.

Así las cosas, sobre la base de todo lo dicho hasta ahora, parece que tenemos tres manera de entender cómo sería la relación entre el juez –el argumentador– y las razones para un caso particular que debe adjudicar:

1) El juez *alikantiano*: el juez acude virgen de decisiones al baúl de las razones y escarba en él para llegar a tomar la decisión; es el buceo en ese baúl lo que permite al juez –simultáneamente– descubrir la decisión correcta y justificarla.

2) El juez *haidtiano*: el juez acude al baúl de las razones provisto de una decisión tomada de antemano y escarba en él para dar con una justificación de la misma, pero es prácticamente imposible que en ese buceo –y por ese buceo– cambie su decisión.

3) El juez *merberiano*: el juez acude al baúl de las razones provisto de una decisión y escarba en él para dar con una justificación de la misma y, si no encuentra una lo suficientemente plausible, ese buceo en el baúl de las razones puede llegar a alterar su decisión hasta dar con una decisión que esté justificada de forma correcta.

4. ¡ES UN IDEAL!

Mi argumento ha consistido en decir que lo que gobierna nuestras vidas morales y normativas no es la *Razón* sino la *razón*. Parece que me muevo, en consecuencia, en el plano descriptivo. Sin embargo, se me podría decir que la interpretación más fuerte de la Escuela alicantina no es en realidad una descripción de la manera en que de hecho razonamos, sino, más bien, una propuesta normativa acerca de cómo deberíamos razonar o, mejor dicho, acerca de cómo deberían razonar los jueces.

Acepto que, por la manera en que he desarrollado hasta ahora el argumento, alguien pueda tener la impresión de que mi crítica está desenfocada: yo estoy diciendo cómo de hecho razonamos, mientras que los alicantinos están interesados en el plano normativo, el ideal del juez argumentador.

Sin embargo, creo que se trata de una objeción errónea.

En primer lugar, es una tendencia bastante generalizada, por lo menos hasta hace relativamente poco, la de proponer grandes ideales en el ámbito de la filosofía práctica normativa (jurídica, moral o política) sin prestar ninguna atención a la evidencia empírica existente respecto de algunas premisas fácticas que esos mismos ideales presuponen, afirman o aceptan. El ideal se trata, se nos dice siempre, de algo a lo que aspirar, algo que debe guiar nuestras prácticas, a sabiendas de que el ideal no será alcanzado porque, de hecho, es imposible.

No hay que ser un naturalista radical para darse cuenta de que esa manera de concebir los ideales regulativos es, en el fondo, de poca ayuda. La Escuela alicantina, en su versión más fuerte, presupone la *Razón*, pero esta última no encuentra evidencia que la apoye: ¿no tiene entonces algo de desquiciado seguir insistiendo en un ideal que presupone algo de lo cual no hay rastro? A mí, el ideal del juez *merberiano* gobernado por la *razón*,

que pone el énfasis de la corrección en el contexto de justificación y no en el de descubrimiento, me parece ya un ideal regulativo muy ambicioso[13].

En segundo lugar, y en la misma línea de lo dicho en el párrafo anterior, lo que pide la versión más fuerte de la argumentación es que los jueces estén motivados por las razones correctas para poder decir que el argumento es, desde el punto de vista material, correcto[14], pero lo que muestra la evidencia es que qué nos motiva es una cuestión no-volitiva, automática e inconsciente[15]. Insistir en controlar las razones que motivan al juez es insistir en algo sobre lo cual no tenemos control. Sobre lo que sí tenemos control –y sobre lo que los propios jueces tienen control–, en cambio, es sobre la manera en que los jueces justifican sus decisiones. Es ahí donde opera la *razón* y donde se mueve el juez *merberiano*.

Un ideal del Estado de derecho empíricamente informado debería exigir a los jueces que justifiquen correctamente las decisiones correctas y que sean sensibles a razones si las justificaciones son incorrectas, lo cual, por lo menos a mí, me parece bastante exigente.

Lo sostenido aquí está sujeto, obviamente, a la evidencia empírica recogida en psicología experimental. Podría ocurrir en el futuro que nueva evidencia empírica contradijera lo dicho aquí. Si eso ocurre, las ideas expresadas en esta sede deberían ser modificadas.

---

13 Por lo demás, diría que al apostar por esta posición estoy reivindicando la concepción estándar de la argumentación judicial. Véase Feteris, 1999: 10; Alexy, 1989: 416, o Bankowski, MacCormick, Summers y Wróblewski, 1991: 17.

14 En realidad, también la versión débil de la Escuela alicantina afirma que los jueces deben estar motivados por las buenas razones. La diferencia es que, para la versión débil, la corrección del argumento material no parece depender de esa actitud. Los buenos jueces –los que están motivados por las razones correctas– parecen ser, en el imaginario de la versión débil de la Escuela alicantina, necesarios para el ideal del Estado de derecho, no para la corrección de la argumentación material.

15 "Mucho más perturbador es el descubrimiento de que incluso en el caso de ser conscientes de cuál parece ser nuestra motivación, nuestros verdaderos motivos pueden ser inconscientes y no ser ni siquiera accesibles a la introspección" (Mercier y Sperber, 2017: 119). Si el proceso introspectivo difícilmente puede revelar qué razones nos motivaron, probablemente será aún más difícil que el derecho, es decir, un tercero, pueda acceder a ellas.

Sea como sea, la evidencia de la que se ha dado cuenta acá acerca de la manera en que razonamos no se refiere al ámbito judicial, sino al del razonamiento moral o al del razonamiento en general. Aunque es cierto que hay diferencias remarcables entre el razonamiento judicial y el razonamiento moral o el razonamiento en general, singularmente las que tienen que ver con el contexto institucional en el que se da el primero, intuyo que la evidencia se aplica en buena medida también a la argumentación judicial, singularmente a la argumentación judicial constitucional. ¿Por qué? No es ningún secreto que los enunciados constitucionales están formuladas de manera más vaga que los enunciados de las leyes infraconstitucionales, por lo que –asumamos *ex hypothesi*– en un sentido los jueces constitucionales gozan de una libertad parecida a la hora de tomar decisiones a la de quienes toman decisiones morales fuera del ámbito del derecho. Sé que esto suena algo imprudente, dicho así, pero una cosa es cierta: no existen –hasta donde yo sé– estudios psicológico-empíricos respecto de la argumentación judicial constitucional en nuestra cultura jurídica[16], así que lo único que se me ocurre que se puede hacer es, con todas las precauciones, trasladar la evidencia empírica que sí existe al ámbito de la adjudicación constitucional.

Conectado con el punto anterior, no parece que la evidencia empírica tenga algo que decir en principio respecto de la plausibilidad o implausibilidad de la tesis de la única respuesta correcta en el caso de la adjudicación constitucional. Yo soy más bien escéptico al respecto –lo cual es compatible con estar convencido de que la tesis de la única respuesta correcta es plausible respecto de los casos infraconstitucionales–, pero lo cierto es que lo único que nos dice la evidencia empírica es que, si hay una respuesta correcta para todo caso constitucional, la *Razón* no es el instrumento evolutivamente conservado para detectarla. Si en la adjudicación constitucional hay una única respuesta correcta, como Mercier y Sperber parecen suponer respecto de la ciencia, esta será normalmente captada por las inferencias intuitivas que representen de manera adecuada el mundo y,

16 Aunque sí en la cultura jurídica estadounidense; véase, por ejemplo, SUNSTEIN, SCHKADE, ELLMAN y SAWICKI, 2006.

en algunas ocasiones, por el intercambio de razones en el entorno social adecuado[17], entorno en el cual explotar la *razón*.

Por supuesto, alguien, quizás afín a alguna forma de constructivismo ético, podría rechazar que las verdades constitucionales estén allá fuera en el mundo, como por hipótesis lo estarían, según alguna forma de realismo científico, las verdades científicas (o por lo menos los hechos que actúan como hacedores de verdad); las verdades constitucionales serían verdades normativas, al fin y al cabo, no reducibles a hechos sociales o hechos naturales. Esto haría que esas verdades constitucionales no pudieran ser aprehendidas por la intuición, dado que el módulo biológico en el que está basada la intuición captura regularidades que ocurren en el mundo natural y social (lo cual no quiere decir que esas regularidades sean irrelevantes para hipotéticas verdades constitucionales, lo que quiere decir es que esas verdades no podrían ser representaciones del mundo natural o social). En tal caso, esa persona debería depositar todas sus esperanzas en el entorno social en que la *razón* debería actuar[18].

---

17 Al respecto, véase la siguiente nota.

18 Haidt –de manera algo incongruente con el resto de su libro– manifiesta: "si juntas a unos individuos de la forma correcta, de manera tal que esos individuos puedan usar sus poderes de razonamiento para desacreditar las afirmaciones de otros individuos, y todos los individuos sienten algún vínculo o destino compartido que les permite interactuar civilmente, puedes crear un grupo que termine produciendo buen razonamiento como una propiedad emergente del sistema social" (2012: 105; trad. propia). Lo que cabe preguntarse es si el derecho es una de esas instituciones en que los participantes tienen un objetivo común de manera tal que interactúan de forma amigable y son sensibles a las razones de los demás. Algunos sostienen que no lo es (como Pierluigi Chiassoni o Diego Moreno Cruz, entre otros), porque no es un juego cooperativo. Si así fuera, entonces el derecho parecería un entorno institucional más bien inidóneo para que prospere la *razón*. No estoy en condiciones de abordar aquí esta discusión. Me limitaré a asumir que el derecho es un juego cooperativo en relación con lo que Raz denomina "el derecho determinado" (trad. propia). Esto dejaría abierta la posibilidad de que en la adjudicación judicial no se tratara de un juego cooperativo. No me parece obvio que sea siempre así, aunque estaría dispuesto a admitir que tal situación es el caso en aquellas disputas de derechos fundamentales que –por hipótesis– admiten una pluralidad de respuestas.

Pero incluso suponiendo el caso contrario, esto es, incluso suponiendo que (por lo menos) los casos difíciles en la adjudicación constitucional no tienen una respuesta clara, "puede haber intuitivamente mucho mejores argumentos para una respuesta y no para otra" (Mercier y Sperber, 2017: 198; trad. propia). El reino de la *razón* no es el reino de la arbitrariedad. En este sentido, "las razones no tienen por qué ser meros sellos de aprobación de nuestras intuiciones de primer orden: a veces pueden fortalecer o debilitar nuestras intuiciones de primer orden, y a veces conducen a su revisión" (ibíd.: 146; trad. propia). Lo único que se pretende desacreditar aquí es que los juicios normativos sean descubiertos por la *Razón*.

Desligándonos por un momento del dominio de lo jurídico, es necesario señalar que del hecho de que estemos evolutivamente conservados no para buscar la verdad moral sino para racionalizar nuestras intuiciones no se sigue que no haya verdad moral –tampoco lo contrario–, solo quiere decir que el razonamiento no está evolutivamente conservado para ello. La teoría de la *razón* es lógicamente independiente de la cuestión metaética.

Recuérdese que la versión débil de la Escuela alicantina de la argumentación parecía afirmar que forma parte del ideal del Estado de derecho –aunque no de aquello que significa un argumento correcto desde el punto de vista material– que el juez tenga la actitud correcta (*i.e.*, que esté motivado por las razones correctas). Sin embargo, el ideal del Estado de derecho es un ideal acerca de un dominio normativo, no psicológico; versa acerca de reglas y criterios normativos, no sobre actitudes psicológicas o emotivas. En la medida en que el juez sepa justificar correctamente sus decisiones correctas normativas no importa mucho, desde el punto de vista estrictamente conceptual, cuáles sean sus actitudes, sus aficiones o incluso sus posibles filiaciones.

Por último, es necesario señalar que, si la *razón* es el caso, entonces el argumento de Aguiló a favor del principio de imparcialidad e independencia judiciales parecería quedar algo desacreditado. Yo creo que la justificación de esos mencionados principios, así como su necesaria presencia en el derecho, no proviene de un argumento de impronta kantiana como el que, con su elegancia habitual, plantea Aguiló. El principio de imparcialidad e independencia judiciales están justificados no en virtud de razones morales o de razones conceptuales, sino en virtud de razones institucionales: la estabilidad y la credibilidad de una institución como el derecho,

y, con más razón, del Estado de derecho, depende, al menos en parte, de eliminar la impresión –o, si se quiere, la apariencia– de parcialidad o de no-independencia de quienes toman las decisiones. Es más fácil confiar en quien te dice que decide imparcialmente si confías en él, y es más fácil confiar en él si eliminas cualquier sospecha de parcialidad. La justificación de esos principios no bebe de Kant –y de su punto conceptual–, sino del sentido institucional de quien diseña las instituciones[19].

CONCLUSIONES

La concepción argumentativa fuerte de la Escuela alicantina es muy ambiciosa. Se apoya en una manera de entender el razonamiento que hace que la argumentación sea un instrumento que nos acerca a la verdad (jurídica). Esta manera de entender el razonamiento parece heredar de Kant una gran confianza en las capacidades de razonar de los humanos. Como consecuencia de ello, la famosa distinción neopositivista entre contexto de justificación y contexto de descubrimiento –así como la paralela entre razones explicativas y razones justificativas– trasladada a la adjudicación quedaría desacreditada.

Sin embargo, la evidencia empírica parece no avalar esa concepción del razonamiento de impronta kantiana y, como consecuencia, haría menos severo el debilitamiento de la distinción entre contexto de descubrimiento y contexto de justificación.

Así las cosas, parece que la manera de determinar cuándo una argumentación en el derecho es correcta –asumiendo la tesis de la única respuesta correcta– reside básicamente en el contexto de justificación. No es tan importante cómo se llega a esa respuesta –entre otras cosas porque no tenemos acceso a ni podemos controlar ese "cómo"–, sino cómo debería justificarse la decisión correcta. Como consecuencia de ello, y contra lo que sostienen los miembros de la Escuela alicantina de la argumentación en su

19 De hecho, como mencioné, el propio AGUILÓ (2009: 30, 32, 35) menciona también razones de este tipo a favor de estos principios, pero lo hace mezclándolas, diría que de manera inadvertida, con el argumento conceptual de impronta kantiana.

versión más radical, la manera más razonable de concebir la argumentación no sería como una actividad, sino como un resultado.

Todas estas conclusiones son provisionales, y este trabajo es solo el primer paso en un campo de investigación que podría explotar en el futuro, a saber, el campo en que se entrecruzan las discusiones de teoría de la argumentación jurídica y de la psicología experimental.

El juez *alikantiano*, en fin, parece suscribir aquella idea de T. S. Eliot según la cual no hay mayor pecado que hacer lo correcto por las razones incorrectas. Pero no es la primera vez que el valor poético de una idea es mayor que su valor científico, filosófico o normativo.

REFERENCIAS

Aguiló, J., 1997: "Independencia e imparcialidad judicial y argumentación jurídica", *Isonomía*, 6, 71-79.

Aguiló, J., 2009: "Imparcialidad y concepciones del derecho", *Jurid. Manizales*, 6 (2), 27-44.

Alexy, R., 1989: *Teoría de la argumentación jurídica*. Madrid: cepc.

Atienza, M., 1999: "El derecho como argumentación", *Isegoría*, 21, 37-47.

Atienza, M., 2014: *Curso de argumentación jurídica*. Madrid: Trotta.

Bankowski, Z., MacCormick, N., Summers, R. y Wróblewski, J., 1991: "On method and methodology", en MacCormick, N. y Summers, R. S. (eds.), *Interpreting Statutes*. London-New York: Routledge, 9-29.

Feteris, Evelyn T., 1999: *Fundamentals of Legal Argumentation. A Survey of Theories on the Justification of Judicial Decisions*, Kluwer, Dordrecht.

Guastini, R., 2014: *Interpretar y argumentar*, Centro de Estudios Políticos y Constitucionales, Madrid.

Haidt, J. y Bjorklund, H., 2008: "Social intuitionists answer six questions about morality", en Sinnott-Armstrong, W. (ed.), *Moral Psychology*, vol. 2, *The Cognitive Science of Morality.* Cambridge, Mass.: mit Press, 181-217.

HAIDT, J., (2012: *The Righteous Mind. Why Good People Are Divided by Politics and Religion*. London: Penguin.

HART, H. L. A., 1972: "Problems of Philosophy of Law", en EDWARDS, P. (ed.), *The Encyclopedia of Philosophy*. Vol. 5/6. New York-London: MacMillan, 261-276.

KENNEDY, D. (1996: "Strategizing strategic behavior in legal interpretation", *Utah Law Review*, 785-825.

KLATT, M., 2008: *Making the Law Explicit*. Oxford: OUP.

MAZZARESE, T., 1995: "Scoperta vs. giustificazione. Una distinzione dubbia in tema di decisioni giudiziali", *Analisi e Diritto*, 145-196.

MERCIER, H. y SPERBER, D., 2011: "Why do humans reason? Arguments for an argumentative theory", *Behavioral and Brain Sciences*, 34, 2, 57-74.

MERCIER, H. y SPERBER, D., 2012: "Reasoning as social competence", en LANDEMORE, H. y ELSTER, J. (eds.), *Collective Wisdom. Principles and Mechanisms*. New York: CUP, 368-393.

MERCIER, H. y SPERBER, D., 2017: *The Enigma of Reason*. Cambridge: HUP.

SCHAUER, F., 2009: *Thinking Like a Lawyer*. Cambridge: HUP.

SUNSTEIN, C., SCHKADE, D., ELLMAN, L. y SAWICKI, A., 2006: *Are Judges Political? An Empirical Analysis of the Federal Judiciary*. Washington D.C.: Brookings Institution Press.

WILSON, T. D., 2002: *Strangers to Ourselves: Discovering The Adaptive Unconscious*. Mass.: HUP.

V.
Acerca de la relevancia de las investigaciones sobre sesgos implícitos para el control de la decisión judicial

FEDERICO JOSÉ ARENA*

INTRODUCCIÓN

La categorización social (*i.e.*, la clasificación de personas en grupos sobre la base de determinadas características [TAJFEL, 1978]) es un componente básico de nuestro modo de pensar y actuar (LAKOFF, 1987). Asimismo, es una herramienta indispensable en el funcionamiento del derecho, puesto que las normas generales asocian una solución normativa a una clase de casos identificada, en parte, mediante referencia a una categoría de personas (ALCHOURRÓN y BULYGIN, 1971). No obstante, el uso de categorías sociales no está exento de dificultades. En ámbito jurídico se exige a los jueces que contrasten los efectos perjudiciales derivados de clasificaciones basadas en características sensibles (o sospechosas), referidas a grupos que han sufrido discriminación en el pasado (SABA, 2009). Numerosos ordenamientos jurídicos, con mayor o menor éxito, han introducido dispositivos para reducir la discriminación producida por categorizaciones sospechosas[1].

Sin embargo, desde hace unos años ha surgido una preocupación por un tipo de discriminación diferente, cuyos efectos no podrían ser combatidos mediante los instrumentos conocidos, puesto que estos últimos resultarían efectivos únicamente contra formas explícitas de discriminación. Esta preocupación se refiere a formas de discriminación producidas por creencias y/o actitudes que, de manera inadvertida para sus portadores, inciden sobre el modo en que clasifican a los demás y reaccionan frente a ellos. En efecto, un cúmulo considerable de trabajos en esos campos han apoyado la tesis según la cual es bastante común que las personas

---

* Doctor europeo en Filosofía del Derecho y Bioética Jurídica de la Universidad de Génova (Italia). Ha sido investigador posdoctoral en la Universidad Bocconi (Milán, Italia). Actualmente es investigador adjunto del Consejo Nacional de Investigaciones Científicas y Técnicas–CONICET (Argentina), profesor adjunto de Filosofía y Lógica Jurídica de la Universidad Blas Pascal (Argentina) y director de la revista *Discusiones*. Agradezco a Flavia Carbonell y a Pau Luque por haber comentado y criticado una versión anterior de este trabajo.

1 Por ejemplo, en Argentina pueden verse, entre otras, INADI (2005) y las leyes nacionales n.º 23.592 sobre medidas para evitar impedimentos arbitrarios al ejercicio de los derechos y garantías constitucionales, n.º 26485 de protección frente a la violencia contra las mujeres y n.º 26618 de matrimonio igualitario.

atribuyan, sin advertirlo, determinados rasgos a los miembros de ciertos grupos y asuman, también sin advertirlo, determinadas actitudes frente a esos grupos y sus miembros (Greenwald y Hamilton Krieger, 2006). Las creencias y actitudes de este tipo son frecuentemente denominadas "sesgos implícitos". Si bien más abajo me referiré a algunas imprecisiones de la expresión, los sesgos implícitos parecen representar un desafío para la lucha contra la discriminación puesto que sugieren la posibilidad de que las personas actúen de manera sesgada y traten a los demás de manera discriminatoria aun sin darse completamente cuenta de ello[2]. Es en virtud de estas posibles consecuencias que algunos teóricos del derecho han promovido la formulación de una agenda pública para lidiar con la incidencia de los sesgos implícitos en distintas áreas institucionales, incluyendo la decisión judicial (Jolls y Sunstein, 2006).

En efecto, investigaciones llevadas a cabo en ámbito jurídico pretenden mostrar que también los funcionarios encargados de aplicar la ley están sometidos a la influencia de sesgos implícitos. La categorización social cumple funciones tanto en la interpretación de los textos normativos como en la identificación de los hechos relevantes del caso. En esos dos ámbitos, los jueces deben llevar a cabo razonamientos que involucran la percepción de sujetos y su inclusión en grupos. Las categorías usadas pueden ser conscientes y, por lo tanto, estar bajo el control del juez; pero también, se afirma, pueden ser reflejo de sesgos implícitos. Así, la categorización puede influir tanto en la justificación externa de la premisa normativa como en la justificación externa de la premisa fáctica. En el primer caso, porque la elección de una determinada interpretación puede haber sido provocada por las propias categorías del juez. Es decir, entre dos interpretaciones posibles, el juez elegirá la que sea consistente o coherente con ellas (Arena, 2016). En el segundo caso, las categorías pueden determinar la manera en

2 En palabras de Jennifer Saul: "Todo esto debería perturbarnos profundamente. Pues significa que nuestros juicios están siendo dramáticamente injustos, aun cuando no lo sean de manera intencional. Tratamos a los miembros de grupos estigmatizados de manera incorrecta, incluso si deseamos con desesperación tratarlos correctamente. Aún más, que todo esto suceda contribuye a que ese trato injusto se perpetúe" (Saul, 2013: 246; trad. propia).

que el juez concibe la explicación de ciertos hechos y, por lo tanto, influir tanto en el grado de plausibilidad otorgado a ciertos elementos probatorios como en el grado de verificación conferido a ciertos enunciados fácticos (Coloma, 2010).

Por ejemplo, respecto de la interpretación, algunas investigaciones han buscado mostrar que los sesgos pueden incidir en el modo en que los jueces asignan significado a las disposiciones normativas. En un caso, se solicitó a un grupo de jueces que evaluaran el significado de una hipotética ley, destinada a despenalizar la tenencia de marihuana cuando existiera una declaración jurada de un médico donde conste que el acusado posee la marihuana por necesidad médica. El problema interpretativo que se proponía era si esa ley cubría también el caso de alguien que, no teniendo la declaración jurada al momento del arresto, la conseguía con posterioridad. De acuerdo con ese estudio, los rasgos del hipotético acusado inciden implícitamente en la interpretación. Así, los jueces interpretaron restrictivamente la ley cuando el acusado tenía alrededor de 19 años y usaba la marihuana para combatir convulsiones, y la interpretaron de manera extensiva cuando tenía 55 años y la usaba para mitigar los dolores de un cáncer de huesos (Wistrich *et al.*, 2015).

Los ejemplos son también numerosos respecto del razonamiento probatorio. El problema suele ser más agudo cuando se trata de delitos contra la integridad sexual. Tan es así que el Estatuto de Roma de la Corte Penal Internacional establece que "[l]a credibilidad, la honorabilidad o la disponibilidad sexual de la víctima o de un testigo no podrán inferirse de la naturaleza sexual del comportamiento anterior o posterior de la víctima o de un testigo", y que "no se admitirá pruebas del comportamiento sexual anterior o ulterior de la víctima o de un testigo"[3].

En todos estos casos, se afirma, la posibilidad de que sesgos implícitos influyan en la decisión del juez aumenta el efecto perjudicial de la discriminación. Por un lado, porque es posible que los jueces decidan de manera discriminatoria aun sin darse completamente cuenta de que lo están haciendo. Por otro lado, porque el impacto discriminador que puede

---

3 Véase Asensio *et al.* (2010: 83-112) para numerosos ejemplos de jurisprudencia argentina en la materia, donde se reflejan tales sesgos en el razonamiento probatorio.

tener el proceso de categorización social es mayor cuando se trata de sesgos implícitos intrainstitucionales. En efecto, si bien los sesgos implícitos de los operadores jurídicos suelen reproducir formas de categorización ya existentes en la sociedad a la que pertenecen, el hecho de que tales sesgos determinen ciertos resultados institucionales puede profundizar (o incluso zanjar) la diferenciación social discriminatoria.

Estas investigaciones parecen apoyar entonces la afirmación de que las decisiones judiciales, como gran parte de otras acciones que involucran la categorización social, pueden sufrir la incidencia de sesgos implícitos. Sin embargo, desde mi punto de vista, queda todavía por precisar exactamente qué rendimiento pueden tener estas investigaciones a los fines de elaborar estrategias de control de la decisión judicial. Es decir, creo que es necesario aún avanzar en esclarecer de qué manera las investigaciones sobre sesgos implícitos mejoran o agregan elementos a los mecanismos ya existentes de control de la decisión judicial. Por ello, en este trabajo quisiera detenerme en el modo en que las investigaciones sobre sesgos permitirían identificar decisiones judiciales equivocadas, concentrándome sobre todo en las decisiones interpretativas, *i.e.*, en las decisiones judiciales acerca del significado de las disposiciones normativas[4]. Esta indagación exige, desde mi punto de vista, abordar preliminarmente algunas dificultades conceptuales.

Una de esas dificultades es cierta ambigüedad en la noción de sesgo implícito, entre sesgo como causa y sesgo como resultado. Para elucidar esta ambigüedad, propondré en el siguiente apartado (1) una mínima cartografía conceptual y terminológica, para asegurar un entendimiento compartido de los términos que serán usados. Ello es necesario ya que, por un lado, hay más formas problemáticas de categorización social que los sesgos implícitos, como los estereotipos y los prejuicios, y, por otro lado, hay más formas de deficiencia cognitiva que la involucrada en la categorización social.

A continuación de ese recorrido conceptual, abordaré dos discusiones suscitadas entre quienes se dedican al tratamiento de los sesgos implícitos tanto a nivel empírico como filosófico. La primera, a la que dedicaré el

4 Si bien haré también algunas referencias a las decisiones en ámbito probatorio.

apartado 2, es acerca de la validez científica de los experimentos que han sido llevados adelante en ciencias cognitivas para mostrar la existencia de sesgos implícitos y sobre cuya base se propone la agenda institucional mencionada más arriba. La segunda, analizada en el apartado 3, es acerca de las conclusiones escépticas sobre cuestiones de epistemología normativa que, se afirma, surgen de las investigaciones sobre sesgos implícitos.

Reconstruiré brevemente los extremos de las dos discusiones mencionadas. Mi intención no es resolver cada una de ellas, sino identificar algunas conclusiones que pueden extraerse a partir del modo en que tales controversias son planteadas. En este sentido, la conclusión más conservadora que propondré es que, incluso si los sesgos implícitos son un componente de nuestro modo de razonar y comportarnos, es necesario distinguir entre, por un lado, el hecho de que la acción sea causada por una actitud o estado mental inconsciente o fuera del control del agente y, por otro lado, el hecho de que el resultado de esa acción sea correcto o incorrecto según cierto criterio normativo. La distinción es importante, no solo por razones terminológicas (es frecuente encontrar en la literatura un uso impreciso del término "sesgo" para hacer referencia, sin distinguir, tanto a la causa como al resultado de la acción) sino, más que nada, porque permite, por una parte, advertir los compromisos conceptuales de las afirmaciones acerca de sesgos implícitos y, por otra parte, precisar la incidencia o relevancia de la investigación sobre sesgos implícitos respecto de la decisión judicial. A la presentación de estas tesis estará dedicado el apartado 4.

Finalmente, en el apartado de conclusiones y consideraciones finales introduciré el problema relativo al tipo de responsabilidad, si es que cabe, que es posible atribuir a quien actúa movido por un sesgo implícito.

1. LA MADEJA CONCEPTUAL

Por lo general, mediante un estereotipo se asocia a una categoría de personas, por el solo hecho de pertenecer a esa categoría, o bien un determinado rasgo o características (estereotipos descriptivos) o bien determinado rol (estereotipos normativos). Es decir, se atribuye un rasgo B o rol C a todos los miembros de una categoría en virtud de que, en cuanto poseen la propiedad A, pertenecen a esa categoría (Oakes *et al.*, 1994). El rasgo o el rol atribuidos pueden ser positivos o negativos (*i.e.*, los cumbieros son

violentos, los asiáticos son buenos en matemáticas, las madres deben ser amas de casa, los hombres deben proveer alimentos). Los estereotipos que atribuyen un rasgo suelen ser analizados como generalizaciones. Ello en cuanto, en el mejor de los casos, se trata de afirmaciones acerca de la existencia de una correlación estadística entre poseer la propiedad adscrita a los miembros de un grupo y el hecho de pertenecer a ese grupo (SCHAUER, 2003 y APPIAH, 2000).

Un prejuicio, en cambio, implica una valoración o actitud respecto de los miembros de cierto grupo que puede estar o no asociada a un estereotipo. Es decir, la actitud positiva o negativa frente a los miembros de un grupo puede depender de considerar que tales personas, en cuanto pertenecen al grupo, poseen una cierta propiedad o estar basada directamente en la pertenencia al grupo.

De este modo, es posible distinguir, por un lado, una reacción cognitiva frente a miembros de un grupo (cuando se estereotipa) y una reacción actitudinal (a partir de prejuicios). Ahora bien, se trata de reacciones típicamente explícitas, en dos sentidos. Por un lado, porque el portador del estereotipo o del prejuicio es consciente de ello y, por otro lado, porque ambos tipos de reacción se manifiestan en el comportamiento. Ahora bien, lo que mostrarían las investigaciones recientes es que ambos tipos de reacciones pueden también ser implícitas, donde implícitas se asocia en la mayoría de los casos a la falta de consciencia por parte del portador y, en consecuencia, a la falta de control de sus efectos. Es decir, es posible que las personas reaccionen de manera estereotipada o prejuiciosa sin advertir que dicha reacción ha sido, precisamente, impulsada por un estereotipo o prejuicio. Un punto importante es entonces precisar el sentido de "implícito" para evitar la ambigüedad apenas señalada relativa a "explícito". El punto que caracteriza a los sesgos implícitos es la inconsciencia y su manifestación en acciones no intencionales, es decir, en acciones no destinadas a manifestar ese estado mental o actitud. Así, un sesgo implícito, tal como una atribución inadvertida de un rasgo o rol a una persona por el mero hecho de pertenecer a un grupo (sesgo asociado a estereotipo) o tal como cierta actitud frente a una persona por el mero hecho de pertenecer a un grupo (sesgo asociado a prejuicio), puede dar lugar a acciones sesgadas. Por ejemplo, la tendencia a favorecer el propio grupo de pertenencia es quizás uno de los sesgos más extendidos. En un test aplicado a jueces de

Minnesota, se les pidió que determinaran la indemnización que correspondía otorgar en un caso hipotético de daño producido por el derrame de líquidos químicos peligrosos en un lago. Cuando el hipotético demandado era de Minnesota, la indemnización era la mitad de la indemnización impuesta cuando el demandado era de un Estado diferente (WISTRICH y RACHLINSKI, 2018: 99).

Ahora bien, los estudios que han investigado la incidencia, en nuestras acciones, de procedimientos mentales inadvertidos, han identificado una amplia variedad de sesgos que no se refieren únicamente a categorías sociales. Se trata, como decía más arriba, de otros tipos de sesgos cognitivos. En este sentido, los sesgos en general constituyen estrategias o recursos que el sujeto aplica de manera inconsciente cuando procesa información. La intervención de procedimientos cognitivos automáticos o inadvertidos es inevitable, pues es imposible que todas las tareas cognitivas de la mente sean conscientes. Sin embargo, no es infrecuente que tales procesos resulten sesgados. Entre los posibles sesgos suelen indicarse, entre otros, los siguientes:

(a) Sesgo de disponibilidad: los sujetos suelen considerar como más probable el acaecimiento de un evento cuando les resulta familiar, es decir, cuando pueden recordar sucesos similares.

(b) Sesgo del anclaje: los sujetos suelen estimar una magnitud a partir de un valor inicial, por lo que, a diferente valor inicial, diferente estimación. Algo parecido sucede con los precios de los autos. El comprador sabe que el precio de lista sirve como punto de inicio para la negociación de una rebaja, pero sabe que el pedido de una reducción tiene un límite. El anclaje también parece afectar el razonamiento de los jueces. En un caso se solicitó a jueces indicar el monto que correspondía otorgar como indemnización en virtud de daños producidos por trato inapropiado en el lugar de trabajo. Algunos jueces recibieron, entre el material probatorio, un testimonio de la víctima donde contaba que había visto en televisión que a una mujer le habían concedido una indemnización de 415.300 dólares por un hecho similar. Los jueces que recibieron ese testimonio concedieron una indemnización promedio de 50.000 dólares, mientras que la indemnización promedio concedida por quienes no lo recibieron fue, en cambio, de 6.000 dólares (WISTRICH y RACHLINSKI, 2018: 93).

(c) Sesgo de la confirmación: los sujetos suelen filtrar información (o evidencia) de manera tal que solo impacta en su conocimiento aquella que confirma su posición inicial[5].

Aquí no me interesa el detalle de estos otros tipos de sesgos, pero hay un punto respecto de ellos que resulta relevante para lo que diré más adelante. En cada uno de estos ejemplos es necesario distinguir dos aspectos que suelen ser confundidos bajo el rótulo de sesgo. Por un lado, el hecho de que la acción haya sido causada por un procedimiento automático (sesgo como causa) y, por otro lado, el hecho de que el resultado de la acción (ya sea una decisión, una nueva creencia, etc.) resulte equivocado (sesgo como resultado). Así, una creencia acerca de una magnitud producida por un sesgo de anclaje bien puede, no obstante su origen, ser correcta (*i.e.*, no sesgada). Lo que, en caso contrario, vuelve equivocada a la creencia resultante del sesgo como causa no es el hecho de haber sido provocada por el sesgo, sino el hecho de que transgrede un criterio, explícito, que permite identificarla como equivocada. Es decir, la identificación misma de una creencia sesgada implica la existencia de un criterio de corrección explícito, independiente del sesgo que la causa. Dicho en otras palabras, es posible que la acción haya sido motivada por un sesgo como causa y, sin embargo, que el resultado no sea sesgado, *i.e.*, si satisface los criterios de corrección. Incluso más, la existencia misma de esos criterios de corrección es indispensable para la identificación de los sesgos como causas. Solo si contamos con criterios para identificar resultados sesgados es posible identificar los sesgos-causa, puesto que estos no se expresan de otro modo (eso es precisamente lo que los vuelve implícitos). En la próxima sección regresaré sobre este punto.

2. LA EPISTEMOLOGÍA DE LO IMPLÍCITO

La preocupación por la incidencia de factores inconscientes en los comportamientos no es ciertamente nueva. Sin embargo, la reciente ola de experimentos en ciencias cognitivas ha conferido cierta legitimidad a estas

---

5 Para una lista más nutrida, véase De la Rosa Rodríguez y Sandoval Navarro, 2016 y Muñoz Aranguren, 2011.

investigaciones. Legitimidad que la epistemología empirista suele negar al psicoanálisis. Además, estas investigaciones recientes tienen también la particularidad de presentar un desafío a estructuras conceptuales tradicionales, como, por ejemplo, la teoría analítica de la acción humana que tradicionalmente ha incluido a la intención como elemento definicional del concepto de acción (Von Wright, 1963). Ello en cuanto, a partir de estas investigaciones se afirma que "los agentes no siempre tienen un control consciente e intencional de los procesos de percepción social, de formación de impresiones y de evaluación que motivan sus acciones" (Greenwald y Hamilton Krieger, 2006: 946; trad. propia).

Las investigaciones que apoyan la tesis de la existencia de sesgos implícitos son numerosas y los experimentos llevados a cabo acumulan ya una gran cantidad de iteraciones y réplicas (ibíd.). Quizás los dos ejemplos siguientes pueden ser ilustrativos de este tipo de investigación. Por un lado, el experimento que llevaron a cabo los psicólogos Peters y Ceci acerca de los sesgos en los procedimientos para la evaluación y publicación de artículos académicos. Por otro lado, el test conocido como sesgo del tirador. En el primer caso, Peters y Ceci enviaron a las más importantes revistas de psicología artículos que ya habían publicado, pero con nombres falsos y sin afiliación a una institución prestigiosa. Solo unas pocas revistas detectaron el plagio, y la gran mayoría (casi el 90%) los rechazaron, señalando serias deficiencias metodológicas (Peters y Ceci, 1982). En el segundo caso, se trata de estudios donde se solicita a los sujetos que disparen (virtualmente) si y solo si ven en las imágenes proyectadas a un sujeto armado. La mayoría de los sujetos tendieron a percibir un arma cuando quien sostenía un objeto ambiguo era una persona negra y un teléfono cuando era una persona blanca (Correll *et al.*, 2007).

Más en general, los experimentos más conocidos y que han adquirido cierta estabilidad son el Test de Asociación Implícita (iat, por sus siglas en inglés) y los estudios denominados del detonador/disparador emocional (*affective-priming*).

El iat es un método destinado a medir la velocidad con la que los sujetos reaccionan frente a estímulos que exigen asociar un par de categorías de personas con adjetivos positivos y negativos. En una versión de este test se exige a los participantes que establezcan, usando el teclado de la computadora, si la imagen de un rostro proyectada en el centro de la pantalla

es un rostro de una persona negra o blanca, y si una palabra proyectada en el centro de la pantalla es una palabra positiva o negativa. Las opciones (blanco/negro-positivo/negativo) están ubicadas, por pares opuestos, en los ángulos superiores izquierdo y derecho de la pantalla. La tecla "E" es usada para ubicar la imagen o palabra del lado izquierdo y la letra "I" para ubicarla del lado derecho. Por lo general se comienza por asociar rostros, se continúa asociando palabras, luego rostros y palabras a blanco/positivo y negro/negativo y luego a blanco/negativo y negro/positivo. El sistema mide el tiempo de cada respuesta.

Así, lo que propone el test es que las diferencias en la velocidad de asociación están vinculadas a un sesgo, en el sentido de que, si implícitamente el sujeto asocia una categoría con un valor positivo, ello se reflejará en que su reacción será más rápida cuando tenga que asociar esa categoría con un valor positivo y será más lenta cuando tenga que asociar una categoría diferente con ese mismo valor. En breve, a mayor velocidad en la reacción, mayor incidencia de un sesgo. Por ejemplo, asociar más rápidamente los rostros y las palabras cuando el par es blanco/positivo y negro/negativo implica poseer un sesgo desfavorable hacia los negros.

El test de *affective-priming* también se basa en la medición de la velocidad de la reacción, pero se diferencia del IAT en que busca activar el sesgo mediante la proyección de una imagen que, si bien impacta en sus sentidos, no es advertida por el sujeto que participa del experimento. Así, si luego de la proyección de esa imagen subliminal el sujeto asocia más rápidamente la imagen subsiguiente (esta sí, percibida conscientemente) con un adjetivo positivo, quiere decir que posee un sesgo positivo respecto de la imagen proyectada de manera subliminal. Y viceversa en el caso de que la asociación se produzca más rápidamente con un adjetivo negativo. Por ejemplo, en algunos experimentos se proyectaban de manera subliminal imágenes de personas blancas y de personas negras e inmediatamente después se solicitaba que se clasificaran ciertos objetos.

Ambos tipos de test pueden ser encontrados y, algunos, realizados en línea, y constan de un altísimo número de réplicas y participantes en línea (véase: https://implicit.harvard.edu/implicit/argentina/). Sin embargo, la validez de tales investigaciones ha dado lugar a una encendida controversia. Aquí me referiré brevemente a la discusión entre, por un lado, Gregory Mitchell y Philip Telock (MITCHELL y TETLOCK, 2006) y, por otro, Samuel

Bagenstos (BAGENSTOS, 2007). Si bien esta controversia gira sobre varios ejes, entre ellos el uso retórico de "ciencia" e "investigación científica", a continuación me concentraré en la crítica que, a partir de un conjunto de estándares sobre la validez de investigaciones científicas, Mitchell y Tetlock dirigen contra los estudios sobre sesgos implícitos, y la respuesta de Bagenstos a esa crítica. Mediante la reconstrucción de la controversia no pretendo, ni estoy capacitado para, zanjar la discusión. Me interesa, en cambio, intentar identificar algunas conclusiones acerca de cómo identificar sesgos implícitos que, si bien son independientes de la resolución de la controversia, son relevantes para el problema que aquí nos interesa.

Mitchell y Tetlock lamentan que los autores que insisten en la necesidad de una agenda pública contra la incidencia de los sesgos implícitos no hayan considerado antes los problemas relativos a las credenciales científicas de las investigaciones en las que apoyan esa agenda[6]. En efecto, según estos autores, existen elementos suficientes para dudar de tales credenciales. Para probar esta afirmación, Mitchell y Tetlock proponen evaluar las investigaciones sobre sesgos implícitos a partir de cuatro estándares de validez científica: (1) Validez de constructo, (2) Validez interna, (3) Validez de la conclusión estadística y (4) Validez externa (MITCHELL y TETLOCK, 2006: 1056).

El primer estándar, de validez del constructo, exige que el objeto inobservable de una investigación haya sido adecuadamente operacionalizado o traducido en una variable que pueda ser manipulada y medida a los fines de testear las hipótesis. El problema con las investigaciones sobre sesgos implícitos es, según los autores, que las variables propuestas como medida de los sesgos son ambiguas. Por ejemplo, el IAT propone como medida de los sesgos la variación en la velocidad de asociación. Sin embargo, esa variable puede ser interpretada como consecuencia de otros procesos alternativos. Además, debería existir cierta convergencia en los resultados de mediciones alternativas del mismo constructo, y ello no ha sido el caso entre IAT y *affective-priming*.

El segundo estándar, de validez interna, exige eliminar explicaciones alternativas que puedan dar cuenta del mismo fenómeno. Es decir, los

---

6 Véase, para una defensa de esta agenda, JOLLS y SUNSTEIN, 2006.

experimentos deberían producirse en contextos donde estén controladas variables alternativas e igualmente explicativas de la acción discriminatoria (Mitchell y Tetlock, 2006: 1032-1033). Los autores señalan una lista de interpretaciones y/o explicaciones alternativas tales como: diferencias en la familiaridad con los distintos grupos sociales, amenaza de estereotipo (*i.e.*, el fenómeno, identificado inicialmente por Claude Steele, que se produce cuando las personas, sabiéndose miembros de un grupo estereotipado, tienden a "confirmar la profecía" encapsulada en el estereotipo [Steele, 2010]), compasión o culpa producidas por la situación de alguno de los grupos involucrados en el experimento, conocimiento del prejuicio de otros o existente en la sociedad, entre otros.

El tercer estándar (de validez de la conclusión estadística), por su parte, requiere que las mediciones estadísticas posean cierta precisión. Sin embargo, mientras que se ha demostrado que muchos factores diferentes inciden en el tiempo de reacción (como la flexibilidad cognitiva, asimetrías respecto de la familiaridad del estímulo, entre otras), los desarrolladores del IAT asumen que en todos los casos indican sesgos frente a los estímulos.

Finalmente, el estándar de validez externa exige que la extensión, a contextos no controlados, de las conclusiones obtenidas en el contexto artificial de laboratorio se apoye en investigaciones aplicadas. Dicho en otras palabras, no está justificado asumir sin más que los resultados de laboratorio permiten predecir comportamientos en el mundo real (Mitchell y Tetlock, 2006: 1034). Por ejemplo, si bien el IAT fue también administrado a jueces en diferentes ocasiones y el resultado promedio fue similar al del resto de las personas que participaron, no siempre el resultado del IAT predijo correctamente la reacción de jueces frente a casos hipotéticos que ejemplificaban los sesgos (Wistrich y Rachlinski, 2018: 101).

Frente a estas críticas, destinadas a cuestionar la validez científica de las investigaciones sobre sesgos implícitos, Samuel Bagenstos defiende la agenda pública dirigida a contrarrestar los efectos de los sesgos implícitos. Su argumento principal es que la crítica, supuestamente epistemológica, de Mitchell y Tetlock refleja en realidad un desacuerdo normativo con los defensores de esa agenda pública. Más precisamente, según Bagenstos,

> … muchas de las críticas de Mitchell y Tetlock a la investigación sobre sesgos implícitos descansan, no tanto en una base científica, sino en un conjunto de

> presuposiciones normativas acerca de qué clase de discriminación debería ser prevenida y castigada por el ordenamiento jurídico. En particular, las críticas se apoyan en una concepción demasiado estrecha, basada en nociones de culpa individual, según la cual el derecho debería prohibir únicamente la discriminación que sea el resultado de un *animus* consciente e irracional […] Sin embargo, quienes promueven el uso del derecho para combatir los sesgos implícitos no asumen esa concepción. Por el contrario, conciben a la discriminación como un problema social que –refleje o no la "culpa" de un individuo discriminador– tiene efectos sistemáticamente dañinos sobre las oportunidades de vida de los miembros de ciertos grupos sociales (BAGENSTOS, 2007: 479-480; trad. propia).

Por ejemplo, como vimos, Mitchell y Tetlock afirman que procesos diferentes a la hostilidad racial pueden producir modificaciones en la velocidad de reacción medida por el IAT. Por su parte, Bagenstos sostiene que esta crítica es en realidad normativa, pues de acuerdo con nuestro autor, y a diferencia de lo que parecen sostener Mitchel y Tetlock, no existe ninguna diferencia relevante, desde una perspectiva anti-discriminación, entre que el resultado del sesgo refleje hostilidad o cualquiera de las otras causas señaladas por estos autores. Lo que importa es que el comportamiento y las actitudes limitan las oportunidades de los miembros de grupos minoritarios. Así, la falta de familiaridad con miembros de grupos minoritarios en ciertos contextos laborales hará que quienes pertenecen a grupos mayoritarios mantengan sus reacciones aversivas, y ello producirá efectos negativos, reflejen o no hostilidad. De este modo, siempre según Bagenstos, tampoco es determinante la supuesta falta de satisfacción del cuarto estándar, pues, por ejemplo, si fuese cierto que las personas en contextos de laboratorio actúan de manera sesgada impulsadas por la amenaza de estereotipo, es probable que esa amenaza también incida fuera del laboratorio, y, en ese caso, lo que cuenta, de nuevo, es que producirán efectos sesgados, afectando los derechos de los miembros de grupos minoritarios. Es esta afectación lo que importa a una política de protección antidiscriminatoria (BAGENSTOS, 2007: 485).

Así, para Bagenstos, la crítica que está detrás del trabajo de Mitchell y Tetlock es en realidad una crítica normativa, a saber, que los defensores de la agenda pública contra los sesgos implícitos no han todavía ofrecido una sólida justificación de la respuesta a actos de discriminación sistémica

y no individuales. El punto sobre el que presiona Bagenstos es ciertamente relevante y, como señalé en la introducción, exige elaborar una teoría acerca de la relación entre sesgos y responsabilidad individual. Pero sobre este último punto solo diré algunas palabras en el último apartado. Lo que aquí me interesa resaltar es que de la misma defensa que Bagenstos hace de la agenda sobre sesgos implícitos resulta posible extraer como conclusión la necesidad de distinguir entre el sesgo como causa y el sesgo como resultado. No solo para evitar caer en una discusión meramente terminológica[7], sino también para identificar qué es lo que criticamos de una acción o creencia cuando afirmamos que es sesgada. Bagenstos es incluso más enfático acerca de que lo verdaderamente relevante no es que las acciones hayan sido producidas por un especial estado mental implícito, sino el hecho de que el resultado de estas sea discriminatorio. Para que esta exigencia normativa de Bagenstos tenga sentido es necesario que poseamos criterios explícitos e independientes que permitan evaluar la corrección o incorrección del resultado de la acción. En el caso que nos interesa, que permita establecer si posee o no efectos discriminatorios.

3. LA DUDA ESCÉPTICA

El segundo problema teórico que quisiera abordar aquí se refiere a cuáles son las consecuencias filosóficas, en ámbito epistemológico, que han de ser extraídas a partir de los resultados de las investigaciones sobre sesgos implícitos. Más específicamente, quisiera analizar el argumento según el cual la demostración de la existencia de sesgos, que no advertimos ni controlamos, pero que influyen en nuestro modo de percibir y reaccionar frente a ciertas categorías de personas, implica inevitablemente una conclusión escéptica general acerca de la validez de nuestro conocimiento.

Me concentraré aquí en la formulación que de tal argumento propone Jennifer Saul (Saul, 2013). Saul sostiene que la evidencia empírica acerca de la existencia de sesgos en el modo en que percibimos a los demás pone en jaque una buena porción de la empresa científica. Gran parte de las afirmaciones científicas se apoyan en el testimonio, es decir, la gran mayoría

7 Sospecha que surge leyendo el trabajo de Bagenstos.

de los enunciados que consideramos verdaderos y justificados se apoyan en el trabajo y las afirmaciones de otras personas. Cuando conocemos a través del testimonio, el conocimiento es adquirido a partir de otros y la adquisición de ese conocimiento exige inevitablemente categorizar a la persona sobre cuyo testimonio nos apoyamos. El problema, señala el argumento escéptico, es que esta última actividad, *i.e.*, la inclusión de otros en categorías, está afectada por la incidencia de sesgos implícitos. Si ello es así, entonces las investigaciones sobre cómo los sesgos afectan nuestro razonamiento ofrece fuertes razones para dudar de la validez del conocimiento que hemos adquirido.

Suelen distinguirse tres ámbitos filosóficos en los que puede incidir el escepticismo. En primer lugar, el escepticismo metafísico acerca de X es aquel que niega que existan hechos del tipo X. En segundo lugar, el escepticismo semántico acerca de un universo de discurso afirma que los enunciados de ese discurso carecen de valor de verdad (o, en algunas versiones, que son todos falsos). En tercer lugar, el escepticismo epistémico acerca de un universo de discurso es el que sostiene la tesis según la cual no es posible establecer el valor de verdad de los enunciados de ese discurso. Saul no está interesada en los dos primeros tipos de escepticismo (además, como señalaré más abajo, está constreñida a negarlos). En efecto, la autora no niega que existan hechos respecto de los cuales nuestros enunciados sobre el mundo empírico puedan ser contrastados. Además, Saul admite que algunos de nuestros enunciados sobre el mundo son verdaderos (*i.e.*, los que afirman la existencia de sesgos). Su escepticismo es, tal como ella misma lo señala, eminentemente epistémico, su tesis es que no podemos estar seguros acerca del valor de verdad que atribuimos a un conjunto de enunciados. Veamos cuáles son estos enunciados y las razones que Saul señala a favor de su conclusión escéptica.

Las investigaciones sobre sesgos implícitos, sostiene Saul, muestran que tenemos razones para dudar de nuestras facultades destinadas a la búsqueda del conocimiento, y ello es así puesto que gran parte de nuestro conocimiento está basado en el testimonio. Los sesgos implícitos inciden de manera inconsciente en el modo en que percibimos y clasificamos a las personas. Dado que el conocimiento basado en el testimonio proviene de otras personas, los sesgos influyen también en el modo en que percibimos y clasificamos a las personas a partir de cuyo testimonio construimos nuestro

conocimiento. En consecuencia, existe un alto riesgo de que los enunciados que creemos apoyados en evidencia (testimonial) se encuentren afectados por estos sesgos y, por lo tanto, no podemos estar seguros (*i.e.*, tenemos buenas razones para dudar) de que efectivamente conozcamos su valor de verdad (Saul, 2013: 244). De acuerdo con Saul, gran parte de nuestro conocimiento lo hemos obtenido de esta manera. Las consecuencias de este argumento son incluso más devastadoras, pues llegados a este punto no podemos siquiera identificar qué porción de ese conocimiento se produjo o no de manera sesgada, puesto que no podemos ya recordar el origen exacto de cada enunciado, *i.e.*, no podemos recordar el testimonio exacto a partir del cual hemos establecido la verdad o falsedad de cada enunciado. Como conclusión deberíamos dudar de un gran número de los enunciados que actualmente consideramos verdaderos y justificados.

Saul sostiene que la forma de escepticismo epistémico que defiende es distinta de otras formas tradicionales o más conocidas del mismo tipo de escepticismo como, por ejemplo, la tesis según la cual no tenemos evidencia que permita descartar la hipótesis de que no somos más que cerebros en un balde. Las diferencias entre este tipo de escepticismo y el de Saul son de fuerza, de alcance y pragmáticas. Con relación a la fuerza, Saul señala que mientras el escepticismo tradicional solo nos confiere razones para *dudar* de la verdad de todos nuestros enunciados sobre el mundo externo, el escepticismo producto de los sesgos implícitos nos ofrece buenas razones para *creer* que no podemos confiar en nuestras capacidades epistémicas. En segundo lugar, mientras el escepticismo del cerebro en el balde ofrece razones para dudar de que *todos* nuestros enunciados acerca del mundo exterior son falsos, el escepticismo de los sesgos implícitos alcanza solo un subconjunto de tales enunciados, *i.e.*, los que involucran el testimonio como fuente de conocimiento. Es decir, mientras el escepticismo tradicional es global, puesto que todos nuestros enunciados pueden tener un valor de verdad distinto del que les atribuimos, el escepticismo de Saul es no global (podemos estar seguros del valor de verdad de algunos enunciados). Sin embargo, dada la difusión que, según la misma Saul, el testimonio tiene en el modo en que adquirimos conocimiento, su escepticismo no sería local (*i.e.*, reducido a un conjunto definido y acotado de enunciados) sino casi global, es decir, tenemos razones para dudar del valor de verdad que hemos atribuido a gran parte de nuestros enunciados acerca del mundo.

Tercero, mientras que los efectos pragmáticos del escepticismo tradicional se agotan una vez que nos levantamos del sillón filosófico, el escepticismo de los sesgos implícitos nos exige ponernos en acción. Es decir, del hecho de que seamos cerebros en un balde no se sigue ninguna razón para cambiar el modo en que vivimos, pues el engaño está tan bien pertrechado que nuestras vidas podrían transcurrir igual en ese caso. En cambio, el escepticismo de los sesgos nos exige actuar para evitar que nuestro modo equivocado de clasificar a las personas incida en lo que creemos conocer y, así, incida en nuestras acciones (SAUL, 2013: 244). Este último punto exige entonces que estemos atentos para evitar que los sesgos puedan incidir en nuestras evaluaciones de los demás; según Saul,

> … el problema empieza a hacerse evidente cuando nos preguntamos a nosotros mismos cuándo deberíamos preocuparnos por la influencia de los sesgos implícitos en nuestros juicios. La respuesta es que deberíamos preocuparnos cada vez que tengamos que considerar una afirmación, un argumento, una sugerencia, una pregunta, etc. de una persona cuyo grupo social de pertenencia estamos en condiciones de reconocer (ibíd.: 250; trad. propia).

Ahora bien, me parece que la argumentación que propone Saul tiene algunos problemas que suelen afectar al escepticismo. Problemas que provienen de lo que llamaré el efecto del grifo abierto. Este efecto suele incidir en los escepticismos, supuestamente no globales, sino casi globales. Ello es así porque este tipo de escepticismo, si bien afirma que gran parte de los enunciados de un ámbito del discurso están afectados por el aguijón escéptico, presupone necesariamente que al menos un enunciado de ese ámbito del discurso está libre del defecto, pues es precisamente sobre ese enunciado que se apoya el argumento escéptico. En este sentido, es conocido el análisis que Dworkin lleva a cabo del escepticismo interno, casi global, respecto de los enunciados morales (DWORKIN, 2011: 23-97). Un ejemplo de este tipo de posición es la que afirma que todos nuestros enunciados morales son falsos puesto que carecemos de libre albedrío. Dworkin señala que este escepticismo, a primera vista global, es necesariamente casi global, pues depende de la verdad de al menos un enunciado moral, a saber, el enunciado contrafáctico según el cual si poseyéramos libre albedrío, tendríamos algunos deberes morales. El escepticismo de Saul está también constreñido

a ser casi global, puesto que su tesis, acerca de la incidencia de los sesgos implícitos en la adquisición de conocimiento basado en testimonio, presupone que somos capaces de adquirir conocimiento acerca de la existencia de sesgos implícitos. Sin embargo, y aquí está el efecto del grifo abierto, las razones por las que hemos de dudar de nuestro conocimiento basado en el testimonio se extienden también al conocimiento sobre la existencia de sesgos implícitos. A continuación intento mostrar por qué ello es así.

Una vez que aceptamos que tenemos fuertes razones para dudar de nuestro conocimiento basado en el testimonio, puesto que la adquisición de ese conocimiento exige la categorización de personas y las investigaciones en ciencias cognitivas muestran que sufrimos de sesgos cuando llevamos a cabo tal categorización, tenemos buenas razones para dudar también de las investigaciones en ciencias cognitivas, puesto que los experimentos de laboratorio llevados cabo por tales científicos incluyen la categorización de personas. Quizás Saul podría argüir que, a diferencia de otros defensores de la existencia de sesgos implícitos, ella no niega que seamos capaces de advertir la existencia de tales sesgos y de intentar evitarlos. Es decir, gracias a los resultados de las investigaciones sobre sesgos, ahora podemos (estamos obligados a) estar atentos al modo en que adquirimos conocimiento a través del testimonio y evitar así la influencia de sesgos. Dicho con otras palabras, el aguijón escéptico afecta a todo el conocimiento anterior a estas investigaciones, pero podemos ser optimistas (o, al menos, no tan pesimistas) acerca del futuro, pues ahora sabemos que hemos de estar alerta. El problema con esta respuesta es que los experimentos llevados a cabo en ciencias cognitivas se realizaron antes de obtener su resultado, es decir, antes de saber que nuestra categorización de las personas está precedida por sesgos. Esto nos pone frente al dilema siguiente: o bien aceptamos que hemos logrado conocimiento (*i.e.*, aceptamos que tenemos buenas razones para considerar como verdaderos ciertos enunciados) antes de saber que sufrimos de sesgos implícitos, pero entonces hemos de aceptar que hay otros enunciados, además de los que identifican sesgos, en los que podemos confiar; o bien defendemos un escepticismo casi global y por lo tanto tenemos buenas razones para dudar que hayamos logrado obtener conocimiento antes de saber que los sesgos implícitos influyen en nuestra categorización, pero entonces tenemos buenas razones incluso para dudar de los resultados de las investigaciones sobre sesgos implícitos.

Al igual que en la sección anterior, me parece que la conclusión que puede ser extraída a partir del recorrido apenas realizado y del dilema al que nos ha llevado el escepticismo de Saul es que la identificación de sesgos implícitos, mediante las investigaciones en ciencias cognitivas, presupone criterios de corrección de enunciados sobre el mundo que no pueden reducirse al criterio simple de que no hayan sido causados por sesgos. Criterios que, por lo tanto, bien podemos usar para evaluar el resultado de otros procedimientos cognoscitivos, *e.g.*, de otros enunciados considerados verdaderos y justificados. Es decir, si queremos evitar el segundo cuerno del dilema (el escepticismo global), debemos aceptar que contamos con criterios para identificar enunciados que ofrecen conocimiento y que, sobre esa base, podemos identificar como sesgado un resultado que se aparte de ese criterio. Dicho con otras palabras, la identificación de sesgos-resultado exige que contemos con criterios de corrección independientes a la existencia misma del sesgo implícito-causa.

4. VOLVIENDO AL ÁMBITO JURÍDICO

Veamos cómo el análisis llevado adelante hasta aquí puede ser de utilidad para abordar la incidencia que las investigaciones sobre sesgos implícitos pueden tener respecto del control de la actividad judicial, en particular, del conjunto de acciones a través de las cuales los jueces determinan la existencia y el contenido del derecho. Hay una limitación inicial a la incidencia que los sesgos poseen en la decisión judicial, y es que la actividad judicial se produce en un contexto de decisión extendido en el tiempo. Es decir, los jueces cuentan con la posibilidad de deliberar con mayor o menor detenimiento acerca de qué decisión tomar. En contextos decisorios de este tipo, la incidencia de los sesgos, si bien no desaparece, tiende a disminuir. En este sentido, los psicólogos suelen distinguir entre dos formas de tomar decisiones, a saber, uno intuitivo o automático y otro deliberativo o reflexivo. Esta diferencia suele, a su vez, expresarse en términos de la distinción entre el Sistema 1 y el Sistema 2, respectivamente. Así, el Sistema 1 sería aquel que involucra la intuición y las emociones, produciendo decisiones más veloces y apoyadas, por lo general, en asociaciones automáticas entre conceptos. El Sistema 2, en cambio, involucra capacidades intelectuales de orden superior y produce decisiones razonadas, apoyadas en la deducción

lógica o el razonamiento analógico. Es por ello que las decisiones a través del Sistema 2 insumen mayor tiempo (véase KAHNEMAN, 2012). Obviamente, como señalan varios autores, esta estricta distinción entre dos tipos de formas de tomar decisiones es, en cierto sentido, una ficción puesto que ambos sistemas suelen entrelazarse al momento de tomar una decisión concreta (WISTRICH y RACHLINSKI, 2018: 90).

Ahora bien, incluso en los casos en que los sesgos afecten decisiones tomadas en contextos donde predomina el Sistema 2 cabe preguntarse qué aporte puede esperarse de las investigaciones sobre sesgos para el control de la decisión judicial. A través del recorrido analítico propuesto en este texto he intentado mostrar la ambigüedad teórica de la noción de sesgo implícito. A veces, con la expresión "sesgo implícito" se hace referencia al estado mental que provoca, de manera inadvertida, una determinada acción, mientras que otras veces la expresión es usada para hacer referencia al resultado de ciertas acciones. Desde mi punto de vista, una vez sacada a la luz esta ambigüedad es posible advertir que la identificación misma de la existencia de un sesgo como causa presupone que contamos con criterios explícitos, e independientes de esa causa, para determinar si la acción (su resultado) es sesgado. Es decir, criterios destinados a identificar cuándo el resultado de la acción se aleja de cierto criterio y, en el caso de la categorización social, cuándo el resultado de la acción es discriminatorio.

En el ámbito de la decisión judicial, tales criterios son ofrecidos por las convenciones o métodos interpretativos y por los estándares de prueba. Esto es, la primera exigencia que se impone a un juez, al momento de identificar el significado de las disposiciones normativas, es que respete las convenciones interpretativas, *i.e.*, que interprete correctamente. Asimismo, se exige al juez que, al momento de evaluar los elementos probatorios, decida según el estándar de prueba correspondiente, ya sea el de más allá de toda duda razonable, el de prueba preponderante, mejor explicación u otro. De este modo, una interpretación o una valoración de la prueba aparecerán sesgadas precisamente cuando se aparten de esos criterios. Si ello es así, entonces también en ámbito judicial la identificación misma del sesgo presupone la existencia de un criterio normativo, en este caso métodos interpretativos y estándares de prueba. Pero si ya contamos con esos criterios para el control de la decisión judicial, entonces la inclusión de los sesgos implícitos no agregaría nada. Dicho con otras palabras, existiendo

criterios de corrección de la decisión judicial, lo que importa es que esos criterios sean respetados. Incluso si la decisión fue motivada por un sesgo como causa, si el resultado no es sesgado, *i.e.*, si satisface los criterios de corrección, entonces no existirá un sesgo como resultado.

Por supuesto, se puede todavía argüir que una decisión jurídicamente correcta puede ser, de todos modos, discriminatoria. Pero incluso en este caso, el criterio para determinar la naturaleza discriminatoria de la decisión, que ahora ya es distinto del criterio de corrección jurídico, será un criterio explícito, idéntico a los criterios para identificar acciones producto de las formas tradicionales de categorización social explícita y problemática, como estereotipos y prejuicios. Lo que estoy intentando señalar es que tanto el criterio de corrección jurídica como un posible criterio de discriminación extra-jurídico son criterios explícitos e independientes de la creencia o actitud que causó la decisión.

El argumento anterior podría ser tachado de formalismo ingenuo y cientifismo desmedido pues los criterios de corrección interpretativa y probatoria no están siempre disponibles y, por lo tanto, las consideraciones precedentes pueden parecer algo simplistas desde el punto de vista de la teoría de la interpretación o de la teoría de la prueba. Enfrentar esta observación exige formular una aclaración. Ciertamente existen ámbitos de indeterminación interpretativa y probatoria, más o menos amplios según el ordenamiento jurídico. El argumento anterior no niega esto, ni necesita hacerlo, simplemente señala que, allí donde esos criterios son determinados, los hallazgos sobre sesgos implícitos no agregan nada a las herramientas provenientes de los análisis de las formas explícitas de categorización social, tales como estereotipos y prejuicios. Ahora bien, ¿qué sucede en aquellos casos de indeterminación? La primera impresión es que la respuesta sería que es precisamente en aquellos casos donde no hay criterios de interpretación claros o estándares de prueba precisos donde la decisión se vuelve discrecional y puede verse afectada por un sesgo y, por lo tanto, producir resultados discriminatorios. En este punto vale la pena introducir una distinción propuesta por Ricardo Caracciolo entre diferentes sentidos en que una decisión puede ser discrecional:

(a) Discreción fuerte: es la situación en la que el ordenamiento jurídico no provee ninguna respuesta al caso que debe ser decidido y por lo tanto

no hay ninguna decisión jurídica correcta que respetar. Por ejemplo, en los casos de laguna normativa.

(b) Discreción reglada: es la situación en la que el ordenamiento jurídico otorga al juez la facultad para tomar una decisión, dentro de un conjunto limitado de decisiones alternativas. Por ejemplo, para la determinación de las penas dentro de un mínimo y un máximo.

(c) Discreción conferida: es la situación en la que el ordenamiento jurídico delega en el juez qué decidir caso por caso (Caracciolo, 2009)[8]. Por ejemplo, cuando el ordenamiento expresamente le indica al juez que decida de manera "razonable" o según su "leal saber y entender".

En estos tres casos, y más allá de que en el caso (b) el juez deba decidir dentro de un marco de opciones, no contamos con criterios interpretativos o probatorios para identificar una decisión correcta e incorrecta. Y, por lo tanto, si no hay decisión correcta desde el punto de vista de algún criterio, entonces ninguna decisión es sesgada, al menos desde el punto de vista de criterios jurídico-interpretativos. Obviamente, se puede todavía argüir que la identificación de la decisión sesgada, dado que no hay criterios jurídicos, depende ahora de criterios que permiten determinar su naturaleza discriminatoria. Pero, de nuevo, se tratará también de un criterio explícito, idéntico a los criterios para identificar acciones producto de las formas tradicionales de categorización social explícita y problemática, como estereotipos y prejuicios.

Todo esto parece llevarnos a un dilema, donde ambos cuernos incluyen la tesis de la irrelevancia de los sesgos implícitos. Es decir, o bien poseemos criterios jurídicos para identificar decisiones sesgadas, pero en ese caso nos alcanza con los criterios jurídicos y la investigación sobre sesgos es irrelevante; o bien no poseemos criterios jurídicos y por lo tanto la identificación de decisiones sesgadas exige recurrir a criterios extra-jurídicos, pero en ese caso se tratará de criterios explícitos que, de nuevo, vuelven irrelevante la cuestión sobre la incidencia de los sesgos implícitos.

En realidad creo que no es complicado escabullirse de este dilema. En efecto, quienes defienden una agenda pública, incluido el control de

---

8 Para otras distinciones entre diferentes formas de discreción, véase Dworkin, 1984 (1978) y Fletcher, 1984.

decisiones judiciales, a partir de las investigaciones sobre sesgos, no están preocupados por la decisión individual, sino por el efecto agregado de decisiones sesgadas. En este sentido, creo que el criterio que surge de esa agenda es la exigencia de igualdad transversal caso a caso. Por ejemplo, cuando se lamenta que un juez, "pocas semanas después de imponer una condena de seis meses por agresión sexual a un estudiante universitario blanco de Stanford, imponga una condena de tres años a un inmigrante mexicano por un delito similar" (WISTRICH y RACHLINSKI, 2018: 88; trad. propia).

En estos casos, siempre hay un criterio normativo externo que permite identificar el resultado sesgado, pero no se trata ya de un criterio relativo a la decisión individual, sino de la exigencia de igualdad entre decisiones. Este tipo de criterio permitiría identificar decisiones sesgadas incluso en aquellos casos en los que no hay criterios jurídicos para hacerlo. Si bien el criterio de igualdad ya alcanzaría para imponer esta exigencia y, por lo tanto, los sesgos implícitos podrían, de nuevo, aparecer como irrelevantes, el punto de estas investigaciones es que ofrecen una explicación y por lo tanto un diagnóstico y un remedio, para los casos en que los resultados agregados de las decisiones judiciales muestren un resultado desigual entre grupos sociales, en especial, cuando perjudican a grupos vulnerables o que han sido discriminados históricamente. Decisiones disparatadas en casos símiles permite sospechar de la incidencia de un sesgo.

CONSIDERACIONES FINALES

Generalmente se considera que la actividad (*i.e.*, las decisiones) de los jueces resulta controlada por mecanismos institucionales específicamente previstos a tal efecto, en particular los distintos recursos procedimentales. Estos recursos permiten que un órgano judicial, superior al que tomó la decisión, controle la corrección de la decisión del tribunal inferior. Se trata, en consecuencia, de un tipo de control vertical. Los límites de este control se manifiestan cuando se trata de órganos judiciales supremos, es decir, órganos cuyas decisiones son finales, en el sentido de que no existe un órgano superior que pueda revisarlas. La relevancia de la investigación sobre sesgos implícitos para el control de la decisión judicial individual está limitada, pero no por ello es irrelevante. Es limitada porque, por un lado, la decisión judicial se realiza en contextos deliberativos y no intuitivos,

más allá del papel que ciertamente le psicología del juez juega al momento de tomar la decisión. Por otro lado, porque, o bien existen criterios de corrección jurídica (para cualquier posición interpretativa, distinta del escepticismo radical, los jueces deciden dentro de cierto marco de criterios de corrección), o bien contamos ya con criterios explícitos para identificar decisiones discriminatorias. En esos casos, la corrección de la decisión estará fijada por estos criterios, independientes del sesgo.

De todas formas, las investigaciones sobre sesgos parecen adquirir relevancia cuando observamos las decisiones judiciales en su conjunto. Es decir, cuando nos preocupamos por el carácter sistemático de la discriminación. Desde este punto de vista, el problema es que decisiones individuales pueden aparecer correctas según criterios de corrección jurídica o criterios de control de la discriminación explícita, y sin embargo ser sesgadas si son consideradas dentro del conjunto de decisiones judiciales y se las compara con otras, o si se tiene en cuenta el resultado agregado de todas ellas. Se trata, sobre todo, de una exigencia de igualdad.

Ello permite advertir que, además del modo tradicional de controlar la actividad judicial mediante la posibilidad de recurrir, se vuelve también relevante un modo alternativo de control tendiente a evitar que los sesgos implícitos lleguen a producir el resultado discriminador, diseñando medidas para reducirlos o eliminarlos (incidiendo durante el proceso de selección de los jueces y, posteriormente, construyendo instancias de formación).

Por último, esta incidencia de los sesgos exige volver la mirada a los mecanismos previstos para el control de la actividad de los jueces individuales. Mecanismos tales como los consejos de la magistratura o jurados de enjuiciamiento. Sin embargo, este último punto exige abordar con detenimiento la cuestión de la responsabilidad de quienes deciden bajo la incidencia de sesgos. Al respecto, los sesgos implícitos representan un desafío para la teoría de la responsabilidad, puesto que no resulta del todo claro si está justificado responsabilizar a los sujetos por ellos. Por un lado, hay quienes sostienen una posición exoneradora, puesto que los sesgos implícitos son resultado de la educación, el contexto cultural y el estrato social de pertenencia, por lo que el agente no solo no puede controlarlos, sino que, además, suele no advertir su incidencia en el propio razonamiento. Si ello es así, las acciones sesgadas carecerían de dos rasgos que suelen asociarse con la responsabilidad: control y consciencia (Lawrence, 1987).

Por otro lado, distintas investigaciones muestran que, si bien el funcionamiento de los sesgos implícitos es inconsciente, de ello no se sigue que sea imposible para el agente advertirlos y, por lo tanto, controlarlos, por lo que resultaría, en cierto grado, responsable (Mavda, 2018). Para avanzar en esta disyuntiva suele proponerse extender la teoría de la responsabilidad por acciones emocionales al abordaje de la responsabilidad por acciones sesgadas. En efecto, el modo en que las emociones influyen en la conducta parece similar al modo en que lo hacen los sesgos implícitos. Así, en ciertas ocasiones las emociones resultan de una intensidad tal que desplazan, en la motivación del agente, el balance de razones a favor o en contra de la acción. Por ello, las acciones emocionales suelen ser consideradas como el resultado de factores causales que exceden el control del agente (Manrique, 2014). Pero esta será una tarea para futuros trabajos.

REFERENCIAS

Alchourrón, C. E. y Bulygin, E., 1971: *Normative Systems.* Wien-New York: Springer.

Appiah, K. A., 2000: "Stereotypes and the shaping of identity", *California Law Review,* 88 (1): 41-53.

Arena, F. J., 2016: "Los estereotipos normativos en la decisión judicial", *Revista de Derecho de la Universidad Austral de Chile,* 29 (1): 51-75.

Asensio, R. et al., 2010: *Discriminación de género en las decisiones judiciales: Justicia penal y violencia de género.* Buenos Aires: Defensoría General de la Nación.

Bagenstos, S., 2007: "Implicit bias, 'science', and antidiscrimination law", *Harvard Law & Policy Review,* 1: 477-493.

Caracciolo, R., 2009: "Discreción, respuesta correcta y función judicial", en *El derecho desde la filosofía. Ensayos.* Madrid: Centro de Estudios Constitucionales, 251-260.

Coloma, R., 2010: "El debate sobre los hechos en los procesos judiciales. ¿Qué inclina la balanza?", en Accatino, D. (ed.), *Formación y valoración de la prueba en el proceso penal.* Santiago de Chile: Abeledo-Perrot, 87-117.

CORRELL, J., et al., 2007: "Across the thin blue line: police officers and racial bias in the decision to shoot", *Journal of Personality and Social Psychology,* 92: 1006-1023.

DE LA ROSA RODRÍGUEZ, P. I. y SANDOVAL NAVARRO, V. D. 2016: "Los sesgos cognitivos y su influjo en la decisión judicial. Aportes de la psicología jurídica a los procesos penales de corte acusatorio", *Derecho Penal y Criminología,* 37 (102): 141-164.

DWORKIN, R., 1984 [1978]: *Los derechos en serio.* GUASTAVINO, M. (trad.). Barcelona: Ariel.

DWORKIN, R., 2011: *Justice for Hedgehogs.* Cambridge-London: Belknap Press of Harvard University Press.

FLETCHER, G. P., 1984: "Some unwise reflections about discretion", *Law and Contemporary Problems,* 47: 269-286.

GREENWALD, A. y HAMILTON KRIEGER, L., 2006: "Implicit bias: scientific foundations", *California Law Review,* 94 (4): 945-967.

JOLLS, C. y SUNSTEIN, C., 2006: "The law of implicit bias", *California Law Review,* 94: 969-996.

KAHNEMAN, D., 2012: *Pensar rápido, pensar despacio.* 3.ª ed. Barcelona: Debate.

LAKOFF, G., 1987: *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind.* Chicago-London: The University of Chicago Press.

LAWRENCE, C. R., 1987: "The Id, the Ego, and equal protection: reckoning with unconscious racism", *Stanford Law Review,* 39 (2): 317-388.

MANRIQUE, M. L., 2014: *Dolo, estados mentales y responsabilidad penal.* México: Fontamara.

MAVDA, A., 2018: "Implicit bias, moods, and moral responsibility", *Pacific Philosophical Quarterly,* 99 (S1): 53-78.

MITCHELL, G. y TETLOCK, P., 2006: "Antidiscrimination law and the perils of mindreading", *Ohio State Law Journal,* 67: 1023-1121.

Muñoz Aranguren, A., 2011: "La influencia de los sesgos cognitivos en las decisiones jurisdiccionales: el factor humano. Una aproximación", *InDret,* 2.

Oakes, P., Haslam, S. A. y Turner, J. C., 1994: *Stereotyping and Social Reality.* Oxford: Blackwell.

Peters, D. y Ceci, S., 1982: "Peer-review practices of psychological journals: the fate of published articles, submitted again", *The Behavioural and Brain Sciences,* 5 (2): 187-255.

Saba, R., 2009: "Igualdad, clases y clasificaciones: ¿qué es lo sospechoso de las categorías sospechosas?", en Gargarella, R. (ed.), *Teoría y crítica del derecho constitucional.* Buenos Aires: Abeledo Perrot, 695-742.

Saul, J., 2013: "Scepticism and implicit bias", *Disputatio,* 5 (37): 243-263.

Schauer, F., 2003: *Profiles, Probabilities and Stereotypes.* Cambridge, Mass.: Harvard University Press.

Steele, C. M., 2010: *Whistling Vivaldi. How Stereotypes Affect Us and what We Can Do.* New York: W.W. Norton & Co.

Tajfel, H., 1978: *Differentiation between Social Groups. Studies in the Social Psychology of Intergroup Relations.* London: Academic Press.

Von Wright, G. H., 1963: *Norm and Action.* London: Routledge and Kegan Paul. Hay traducción castellana de P. García Guerrero: *Norma y acción. Una investigación lógica.* Madrid: Tecnos, 1979.

Wistrich, A. y Rachlinski, J., 2018: "Implicit bias in judicial decision making. How it affects judgment and what judges can do about it", en Redfield, S. (ed.), *Enhancing Justice: Reducing Bias.* New York: American Bar Association.

Wistrich, A., Rachlinski, J. y Guthrie, C., 2015: "Heart versus head: do judges follow the law or follow their feelings", *Texas Law Review,* 93: 855-923.

VI.
Los sesgos cognitivos y la legitimidad racional de las decisiones judiciales[**]

ANDRÉS PÁEZ[*]

INTRODUCCIÓN

Durante los últimos años han aparecido muchos estudios que presentan de manera clara y contundente el efecto negativo de los sesgos cognitivos en los procesos judiciales[1]. Desde el momento en que comienza una investigación hasta el instante de la imposición de la pena existen múltiples ocasiones en que los sesgos de los diferentes agentes judiciales, y las heurísticas que utilizan, interfieren negativamente en sus razonamientos y decisiones. No se trata únicamente de sesgos raciales, sexuales, sociales o religiosos, los cuales han sido ampliamente estudiados en el derecho. Se trata de sesgos implícitos propios del procesamiento de información, tanto sensorial como conceptual, que afectan a todos los seres humanos en todos los ámbitos de la vida, incluso cuando toman decisiones difíciles e importantes (CONLISK, 1996), incluyendo las decisiones judiciales (GUTHRIE *et al.*, 2001).

Esta evidencia empírica contrasta fuertemente con el anhelo de que las decisiones judiciales sean el resultado exclusivo de un razonamiento lógico-jurídico correcto. Más grave aún, estos estudios ponen en duda la posibilidad misma de que los agentes judiciales logren ejercer una racionalidad plena dado el efecto sistemático, recalcitrante y porfiado de los sesgos cognitivos. En efecto, existen muy pocas medidas terapéuticas que hayan

---

* Este artículo hace parte del proyecto "Los sesgos cognitivos en el Derecho", financiado por la Facultad de Ciencias Sociales de la Universidad de los Andes.

** Profesor asociado del Departamento de Filosofía y del Centro de Investigación y Formación en Inteligencia Artificial (CinfonIA) de la Universidad de los Andes (Bogotá, Colombia), director del Grupo de Investigación en Lógica, Epistemología y Filosofía de la Ciencia (Philogica) de la misma universidad y coordinador de la Red Latinoamericana de Epistemología Jurídica.

1 Para una introducción al problema, véase SAKS y KIDD, 1980; WEINSTEIN, 2003; MUÑOZ, 2011; PEER y GAMLIEL, 2013. Algunos autores hispanoamericanos (por ejemplo, SORIA y SÁIZ, 2010; DE LA ROSA y SANDOVAL, 2016) utilizan el término "sesgos cognitivos" en un sentido más laxo que incluye la personalidad del juez y su contexto social, político y religioso como elementos que afectan las decisiones judiciales. En este artículo se tomará el término en un sentido más restringido, tal y como es utilizado dentro de la tradición de sesgos y heurísticas en psicología (KAHNEMAN, SLOVIC y TVERSKY, 1982).

demostrado tener un efecto preventivo sobre los investigadores y decisores judiciales, eliminando o disminuyendo la presencia de estos sesgos.

Frente a estas dificultades, una posible estrategia para disminuir el efecto nocivo de los sesgos es enfocarnos, no en modificar el comportamiento de los agentes judiciales, sino en moldear su entorno para que los sesgos tengan menos oportunidades de aparecer. En particular, considero que es más provechoso concentrar nuestros esfuerzos en identificar aquellas reglas procesales que facilitan o limitan el efecto de los sesgos. Desde la manera de identificar a los sospechosos de un delito hasta el orden en que se presentan los testimonios en un juicio, cada detalle estructural y contextual contribuye a propiciar o a impedir que la información utilizada lleve a conclusiones erróneas. Si uno de los propósitos de la epistemología jurídica es "proponer cambios en las reglas existentes para eliminar o modificar aquellas que resulten ser obstáculos significativos para llegar a la verdad" (LAUDAN, 2006: 3), es innegable que una de sus tareas principales debe ser examinar y proponer modificaciones a las reglas procesales con miras a minimizar la probabilidad de que las decisiones judiciales se vean afectadas por dichos sesgos. Las reglas procesales están diseñadas con base en sus efectos sobre el comportamiento y las percepciones de las partes, y es allí donde debemos concentrar nuestros esfuerzos para mejorar la calidad y, en última instancia, la racionalidad de las decisiones judiciales.

En este capítulo quiero presentar inicialmente un breve panorama del problema de los sesgos cognitivos en los procesos judiciales y discutir algunas de las soluciones que se ha intentado implementar, sobre todo en el derecho estadounidense. De esto se ocuparán las secciones 1 y 2. Pero la pregunta de fondo que quiero plantear no se refiere a ningún sesgo cognitivo en particular ni a ninguna regla procesal específica, sino más bien a las implicaciones que tiene el fenómeno de los sesgos cognitivos para la legitimidad de las decisiones judiciales, las cuales deben estar racionalmente fundamentadas. Tras analizar varias posibles formas de entender la racionalidad jurídica esbozaré el que considero es el camino más propicio para restaurar su legitimidad racional. De ello se ocuparán las secciones 3 y 4.

1. LOS SESGOS COGNITIVOS

El estudio de los sesgos cognitivos en el derecho cuenta con una larga y respetable historia. Desde el impacto que puede tener la apariencia física de un testigo hasta la tendencia a encontrar excesiva coherencia en los hechos del caso, todos los momentos de un proceso judicial son vulnerables a las deformaciones introducidas por los sesgos cognitivos. En esta sección voy a introducir algunos de los sesgos más estudiados a modo de ejemplo y explicaré el tipo de evidencia que se ha utilizado para identificarlos en el ámbito del derecho.

Antes de presentar algunos ejemplos de sesgos cognitivos es conveniente explicar brevemente qué son exactamente[2]. Un sesgo cognitivo es una desviación sistemática, involuntaria e inconsciente de una norma o de un estándar de racionalidad al emitir un juicio perceptual o conceptual, al recordar un evento o al hacer una predicción. No se trata de un simple error, sino de comportamientos que ocurren consistentemente bajo circunstancias similares, y que por lo tanto son predecibles y replicables. Son causados principalmente por el uso de heurísticas, atajos que utilizamos en el procesamiento de información, pero también son producto de las limitaciones de nuestra capacidad cerebral de procesar información, de influencias emocionales, morales y sociales, y de distorsiones en el proceso de almacenamiento y búsqueda de información en la memoria.

A continuación presentaré cuatro de los sesgos más discutidos en la literatura jurídica. Existen muchos otros que se podrían incluir, pero mi objetivo aquí no es ofrecer un tratamiento comprehensivo del tema, sino más bien plantear el problema en términos generales con el objeto de poder discutir más adelante sus implicaciones para la legitimidad racional de las decisiones judiciales.

El primer sesgo que quiero discutir es quizás el más importante, no solo en el derecho, sino en cualquier razonamiento de índole inductiva en

---

2 El concepto de sesgo cognitivo fue introducido por los psicólogos israelíes Kahneman y Tversky a comienzos de los años setenta del siglo pasado. KAHNEMAN, SLOVIC y TVERSKY, 1982, ofrece un tratamiento comprehensivo del tema.

la ciencia o en la vida cotidiana. Se trata del sesgo de confirmación[3]. Al evaluar una hipótesis favorecida por un agente, este mostrará una fuerte tendencia a buscar, recordar y darle más valor a la evidencia que confirme esa hipótesis. Por ejemplo, si un juez se forma rápidamente la idea de que el imputado es culpable, su creencia deformará su valoración de las pruebas presentadas durante el juicio. El sesgo es potenciado por factores emocionales como la rabia (Ask y Pina, 2011). Así, si un detective indignado cree firmemente que un sospechoso es culpable de acceso carnal violento, el sesgo de confirmación lo llevará a privilegiar fuertemente la evidencia que confirme la culpabilidad de esa persona, desechando otras posibilidades. Existe evidencia de que los miembros de los jurados, en los países donde existe esta figura, adoptan muy rápidamente una posición acerca de la culpabilidad o responsabilidad del imputado. En experimentos con jurados simulados, la evidencia muestra que el sesgo de confirmación lleva a una rápida polarización en sus posiciones (Myers y Lamm, 1976). Una consecuencia adicional de este sesgo es el exceso de confianza que genera en las decisiones que se toman con base en la evidencia confirmatoria.

Un segundo sesgo que ha sido objeto de numerosos estudios es el sesgo retrospectivo (Fischhoff, 1975). Cuando se tiene conocimiento de un hecho, existe la tendencia a pensar que este era mucho más predecible de lo que en realidad era posible inferir en el momento con la información disponible. En términos más técnicos, es la tendencia a sobreestimar *ex post* las probabilidades *ex ante*. El resultado es que los jueces tienden a asignarle mayor responsabilidad de la debida a un imputado por negligencia. El ejemplo más infame del sesgo retrospectivo es el de una corte de Nueva Jersey que condenó a los ejecutores de un testamento en 1931 por no haber previsto el desplome de la bolsa en 1929 (*In re Chamberlain's Estate*, 1931). Hoy en día en Estados Unidos los corredores de bolsa y las personas que toman decisiones financieras no pueden ser demandadas por los inversionistas si los resultados no son los esperados, pues se considera que cualquier sentencia judicial al respecto necesariamente va a ser afectada por el sesgo retrospectivo (Rachlinski, 1998: 619).

---

3 Aunque a lo largo de la historia varios filósofos habían señalado la existencia de este sesgo, el término fue acuñado por el psicólogo Peter Wason (Wason, 1960).

El sesgo es muy común en casos de negligencia médica y de responsabilidad por fallas en el desempeño de un producto. También ocurre en las disputas sobre patentes. Cuando hay una disputa acerca de los derechos de autor de un invento, se puede caer en la tentación de pensar que no es sorprendente que diferentes personas hayan llegado a él de manera independiente. En el caso *Graham v. John Deere Co.* (1966), la Corte Suprema de Estados Unidos previno justamente acerca del sesgo retrospectivo y aclaró el sentido en el que un invento patentado puede ser obvio y evidente. Finalmente, en los casos de negligencia se ha detectado que la gravedad del daño, y tener conocimiento de que el demandado intentó remediar el daño causado, aumentan dramáticamente la probabilidad de que el sesgo influya en la decisión (LaBine y LaBine, 1996).

Un tercer sesgo es el resultado de la llamada heurística de representatividad (Kahneman y Tversky, 1973). En la vida diaria a menudo debemos decidir a cuál categoría pertenece un evento o una persona. Es muy frecuente que en el proceso de categorización nos dejemos guiar por alguna propiedad sobresaliente y superficial de los eventos o las personas, saltando rápidamente a la conclusión de que estos pertenecen a una categoría sin tener en cuenta evidencia que puede apuntar en la dirección opuesta. Veamos un ejemplo adaptado de uno originalmente propuesto por Tversky y Kahneman (1982). Considérese la siguiente descripción: "José es una persona que siempre está dispuesta a ayudar. Le encantan los niños y es muy organizado y sensible". Si asumimos que José es colombiano, ¿qué es más probable, que José sea un chofer de taxi o un profesor de preescolar? La mayoría de la gente responderá que es más probable que sea un profesor de preescolar porque se concentra en características como la sensibilidad y el amor por los niños como rasgos representativos de un profesor de preescolar. La persona que escoge esta opción está usando la heurística de representatividad. El problema es que la respuesta está equivocada porque ignora la *tasa base* de hombres colombianos que trabajan como taxistas, que es muy alta, y la de hombres colombianos que trabajan como profesores de preescolar, que es extremadamente baja. Con la información proporcionada, la respuesta es que es mucho más probable que José sea un chofer de taxi.

En el derecho, la heurística de representatividad tiene serias consecuencias. La más notoria de ellas es la interpretación errónea de la evidencia

forense, sobre todo cuando esta tiene un carácter probabilístico. Supongamos que en la escena de un crimen se encuentran fibras de un tapete que corresponde al mismo tipo de tapete que hay en la residencia del principal sospechoso del crimen. Existe la tentación de asignarle un valor probatorio muy alto a este hecho, pero si el tipo de tapete en casa del sospechoso es extremadamente común, la evidencia tendrá un valor probatorio nulo. De nuevo, la clave está en la tasa base y no en la similitud entre las fibras. O consideremos un uso mucho más problemático de la heurística de representatividad. En los casos de abuso sexual a menores, los peritos, utilizando la heurística de representatividad, suelen testificar que el comportamiento de un menor "es consistente con" el comportamiento de un menor abusado. La frase tiene la intención de indicar que el comportamiento del menor también se ha identificado en el de otros menores que han sufrido abuso, y el perito pretende que la semejanza sirva de prueba de dicho abuso. Sin embargo, es claro que muchos niños que no han sido abusados pueden presentar comportamientos semejantes a los de niños que sí lo han sido. La coincidencia entre el comportamiento del niño que está siendo examinado y el de otros niños en la categoría de niños abusados es irrelevante a menos que sepamos la frecuencia de este mismo comportamiento en niños que no han sido abusados (Lyon y Koehler, 1996). En términos más técnicos, si consideramos una evidencia E y una hipótesis H, el resultado de dividir la probabilidad de que se dé la evidencia E dada la hipótesis H, por la probabilidad de que se dé la evidencia E dada la falsedad de la hipótesis H, $p(E|H) / p(E|\neg H)$, debe ser significativamente mayor a 1 para que la evidencia E tenga un valor probatorio[4]. En el caso de abuso a menores, si el comportamiento en cuestión se presenta igualmente en niños abusados (H) y no abusados ($\neg$H), entonces E no tiene valor probatorio porque $p(E|H) / p(E|\neg H) = 1$.

En general, al usar la heurística de representatividad los juzgadores tienden a ignorar la tasa base, a desestimar la posibilidad de que haya

---

4 La proporción $p(E|H) / p(E|\neg H)$ es conocida en inglés como el *likelihood ratio* y hace parte integral del teorema de Bayes. Fue sugerida por Lempert (1977) y Lyon y Koehler (1996) como una medida de la relevancia de una pieza de evidencia frente a una hipótesis dada.

semejanzas debidas al azar, y a olvidar que siempre se deben tener en cuenta posibles escenarios exculpatorios como la contaminación o manipulación de una muestra en un laboratorio.

El último ejemplo que consideraré es el del efecto ancla (GUTHRIE, RACHLINSKI y WISTRICH, 2001). Al estimar una fecha o una cantidad, la exposición previa a un número cualquiera, absurdo o no, afecta la estimación. El ejemplo más cotidiano ocurre cuando intentamos establecer el precio de una vivienda. Los corredores de finca raíz intentarán plantar un ancla alto a partir del cual se negocie el precio, y el comprador limitará sus ofertas a un cierto porcentaje por debajo de esa ancla. En el derecho el efecto ocurre especialmente cuando un juez debe decidir la compensación monetaria a un demandante en un juicio de responsabilidad, o cuando debe decidir la duración de una pena privativa de la libertad. En el primer caso, el abogado del demandante exigirá una cantidad muy alta con el fin de anclar la disputa monetaria alrededor de esa cifra, y en el segundo, el fiscal pedirá una pena prolongada esperando conseguir una que se acerque lo más posible a la solicitada. En un estudio realizado en la Comunidad Autónoma Gallega se analizó el efecto ancla en 555 sentencias entre los años 1980 a 1995 con respecto a la imposición de la pena, y se encontró que en mayor o menor medida el 60% de las sentencias estaban guiadas por la solicitud de la fiscalía (FARINA *et al.*, 2002).

Además de estos sesgos, se han estudiado especialmente: el efecto halo, en el cual la apariencia o el estatus social o profesional de un testigo le añade o le resta credibilidad a su testimonio; el efecto marco, en el que la manera en que se describe una situación o se plantea una pregunta afecta la valoración de lo que está en juego; y la criptomnesia, en la cual una idea adventicia almacenada en la memoria no se experimenta como un recuerdo sino como una idea nueva, original y propia. La lista de sesgos y sus posibles efectos en el derecho es bastante larga, pero considero que este breve resumen de los más prominentes nos permitirá plantear el problema de fondo sobre la legitimidad racional de las decisiones judiciales.

2. ¿ES POSIBLE CONTRARRESTAR LOS SESGOS?

La calidad de un sistema judicial depende en gran medida de la calidad de las decisiones que toman sus jueces y de la capacidad de autoevaluación del

sistema. Si existen razones objetivas, basadas en evidencia empírica, para considerar que las decisiones de los jueces presentan errores sistemáticos y frecuentes, un sistema debería tener la capacidad institucional de corregir dichas fallas. Desafortunadamente dicha capacidad ha brillado por su ausencia. Desde hace varias décadas los psicólogos han expresado su frustración ante la indiferencia del derecho frente a los descubrimientos de las ciencias cognitivas (Saks y Baron, 1980; Monahan y Walker, 1988; Small, 1993). Solo a comienzos de este siglo algunas jurisdicciones en Estados Unidos comenzaron lentamente a introducir reformas, por ejemplo, cambiando algunas leyes referentes a la manera en que se presentan los créditos de consumo y revisando las doctrinas de responsabilidad fiduciaria (Persad, 2014). A un nivel más teórico, la influencia de la psicología experimental ha puesto en duda doctrinas fundamentales del derecho anglosajón como el *stare decisis*. La evidencia cognitiva indica que los seres humanos tenemos la tendencia a preferir el *statu quo*, a sobrevalorar las opciones que se presentan por defecto, a seguir la autoridad de otros y a forzar la coherencia de los casos nuevos dentro de patrones bien conocidos (Jois, 2009). Así, el fundamento del *stare decisis* como heurística recaería más en tendencias que reflejan sesgos cognitivos de los seres humanos que en razones jurídicas teóricamente fundamentadas.

A pesar de la creciente influencia de la psicología, el diagnóstico presentado en la sección anterior solo ha llevado a la modificación de un número muy reducido de reglas procesales. En gran medida esto se debe, según Posner (1999), a que las cortes no sienten ninguna presión para reformar y mejorar la calidad de sus procesos de decisión. Además, en la mayoría de los casos los jueces no reciben retroalimentación alguna acerca de la calidad de sus razonamientos, sin contar con que, en algunos sistemas, los que toman las decisiones de primera instancia son jurados que no deben justificar sus actos.

Muñoz Aranguren sugiere que exigir que los jueces y magistrados justifiquen sus decisiones puede ser un antídoto contra la aparición de sesgos:

> La motivación de las resoluciones judiciales, al exigir una exteriorización razonada y razonable de los motivos que fundamentan el sentido del fallo, supone un freno a los impulsos *decisionistas* del juzgador, *producto de meras conjeturas subjetivas y de sesgos cognitivos*, permitiendo de igual forma la

> revisión por parte del órgano jurisdiccional superior del correcto razonamiento lógico-jurídico de la sentencia (2011: 27; cursiva en el original).

Pero esta sugerencia difícilmente solucionará el problema. En primer lugar, el razonamiento *ex post* presentado en la sentencia no necesariamente describe fielmente el razonamiento real que se llevó a cabo en cada etapa del proceso. Es bien sabido que en muchos casos los jueces deciden y después encuentran la justificación de sus decisiones. En segundo lugar, la exposición de razones en las decisiones judiciales muchas veces no deja muy claro cómo se llega a una decisión y se limita a copiar apartes de los escritos de acusación y defensa. Por lo tanto, es difícil determinar si el razonamiento fue afectado por sesgos cognitivos. En tercer lugar, es implausible que las cortes superiores decidan aumentar su carga de trabajo para revisar el razonamiento de los jueces sin que medie una apelación, especialmente cuando dicha revisión probablemente no alterará en absoluto la decisión tomada durante el juicio. Y finalmente, esta sugerencia asume que los integrantes del órgano superior harán una evaluación de los hechos y del razonamiento del juez que sea inmune a los sesgos que puedan estar presentes en la resolución original. No hay razón alguna para pensar que, si un juez ignoró la tasa base, por ejemplo, un magistrado de un órgano superior no vaya a hacer lo mismo. Y si el magistrado del órgano superior ha recibido entrenamiento para no cometer ese error, la solución real sería ofrecerles el mismo entrenamiento a todos los jueces.

A este respecto vale la pena mencionar la propuesta de Amalia Amaya (2013, 2015), según la cual los jueces deben aspirar a comportarse de una manera epistémicamente responsable e intelectualmente virtuosa. Una de sus obligaciones epistémicas, considero yo, es tomar todas las medidas necesarias para evitar los errores generados por los sesgos cognitivos. La idea de que un juez debe ser consciente de sus limitaciones se origina, por supuesto, en la obra de Jerome Frank, quien afirma en el prefacio de 1948 a *Law and the Modern Mind* que "un juez responsable intentará, en la medida de lo posible, llegar a ser consciente de sus sesgos [de carácter individual] y anular su efecto a través de este autoconocimiento" (1949: xx; trad. propia). Como veremos en esta sección, la virtud intelectual, la responsabilidad epistémica y el autoconocimiento serán insuficientes para este propósito. A continuación examinaré brevemente algunos de los intentos que se han

hecho por eliminar o atenuar el efecto de los sesgos cognitivos dentro del derecho. Me limitaré a los sesgos descritos en la sección anterior.

Es difícil identificar cuáles reglas procesales promueven o atenúan la aparición del sesgo de confirmación. Es una tendencia tan arraigada en todos los seres humanos que es difícil imaginar que la simple modificación de una regla procesal pueda disminuir su presencia. Sin embargo, en la literatura psicológica existen algunos remedios metodológicos que han tenido un éxito limitado. La principal estrategia es cuestionar permanentemente los supuestos con los que razonamos, buscar activamente evidencia que no esté de acuerdo con nuestras creencias preestablecidas y no asumir una actitud puramente reactiva ante la información que se nos presenta. Si bien esta estrategia podría funcionar en un sistema inquisitivo, en los sistemas adversariales el juez o los jurados parecen condenados a reaccionar ante la información provista por cada una de las partes. Sin embargo, aunque el juez no pueda buscar proactivamente nuevas pruebas, sí puede imaginar hipótesis alternativas que también expliquen los hechos. Una vez se haya planteado posibles alternativas, puede buscar activamente, en el acervo probatorio, información que pueda confirmar teorías del caso no contempladas por las partes. Obviamente esta es una estrategia de alcance limitado que depende de la capacidad imaginativa del juez y de la existencia de un acervo probatorio considerable. En últimas, el único recurso universalmente utilizable es la educación de los jueces para que sean conscientes de la existencia del sesgo de confirmación y de todas las formas en que se manifiesta, sin que ello garantice en absoluto su no aparición.

Se podría pensar que las decisiones tomadas colegiadamente tienden a disminuir el efecto del sesgo de confirmación, puesto que cada uno de los miembros del tribunal o del jurado se puede ver enfrentado a puntos de vista opuestos a los suyos. Sin embargo, las decisiones grupales dan pie a otra familia de sesgos que no garantizan que la decisión sea más ajustada a los hechos. El sesgo grupal más común en el derecho es el sesgo de polarización. Muchas veces los grupos toman conjuntamente decisiones más extremas que las que defendía inicialmente cada uno de sus miembros. También se ha determinado que los grupos tienden a amplificar los errores de sus miembros individuales, y son presa del efecto cascada, en el que los miembros tienden a converger hacia las opiniones de los que hablaron o

actuaron primero (Sunstein, 2007). El remedio, por tanto, resulta peor que la enfermedad.

El sesgo retrospectivo, por su parte, es uno de los más recalcitrantes y difíciles de neutralizar. Incluso después de saber acerca de su existencia y de tener la opción de recibir un estímulo económico en caso de acertar, el juzgador promedio recaerá en la sobreestimación de probabilidades. El único recurso conocido para contrarrestar sus efectos es la construcción de árboles de fallas como los utilizados en ingeniería. Los árboles de fallas son diagramas que presentan todas las posibles causas que contribuyen a un resultado negativo y sirven de base para hacer un análisis probabilístico de la contribución de los eventos en cada rama (Fischhoff, Slovic y Lichtenstein, 1978). Una metodología tan invasiva es inviable en un sistema de jurados, pero sería posible entrenar a los jueces en su uso. Naturalmente esta solución nos obliga a contemplar la opción de tener jueces expertos y jueces especializados en asuntos técnicos, posibilidad que ha sido ampliamente discutida en la filosofía del derecho (Di Lello, 1993; Oldfather, 2012). Otra posible solución que ha sido contemplada ha sido la de elevar el estándar de prueba en aquellos casos en los que el sesgo retrospectivo puede beneficiar al demandante (Hastie *et al.*, 1999). Esta solución implicaría abandonar el tradicional estándar de preponderancia de la prueba en los casos civiles. Además, como anota Rachlinski, sería una solución que atacaría los síntomas, pero no la enfermedad (Rachlinski, 2000: 71). La solución que han adoptado varias jurisdicciones es la de suprimir evidencia que no fuera conocida en el momento del supuesto acto negligente. Por ejemplo, la regla 407 de las *Federal Rules of Evidence* (FRE) prohíbe la inclusión de información acerca de cómo el demandado intentó remediar el daño que resultó de su supuesta negligencia o del error de diseño de su producto, información que puede afectar la valoración de la probabilidad *ex ante*. Naturalmente esto solo tiene sentido en un sistema de jurados en el que el juez tiene la potestad de filtrar la información conocida por el jurado. Pero en los sistemas continentales este tipo de información será conocida necesariamente por el juez y afectará sus juicios. Finalmente, una forma de neutralizar el sesgo y debilitar una acusación de negligencia es apelar a criterios más objetivos, como el hecho de que el demandado se ajustó a las prácticas acostumbradas o a los estándares de seguridad generalmente utilizados en circunstancias similares.

En el caso del sesgo de representatividad, el olvido de la tasa base es uno de los problemas más recalcitrantes, no solo en el derecho, sino también en personas que tienen que manejar estadísticas en su vida profesional. Una de las salvaguardas contra la lectura equivocada de las pruebas periciales en los sistemas adversariales es que una de las partes tiene la posibilidad de poner en duda la interpretación dada por la contraparte. Más aún, el argumento se puede resolver apelando únicamente a la teoría de la probabilidad sin tener que apelar a hechos en disputa. En los sistemas no adversariales, sin embargo, el juez tiene que interpretar por sí mismo la información que le proporciona el perito. A menos que tenga un entrenamiento mínimo en estadística para ver la importancia de la tasa base en la interpretación de cualquier correlación o categorización, no es difícil que termine utilizando la heurística de representatividad.

Finalmente, el efecto ancla ha demostrado ser igual de recalcitrante que el sesgo retrospectivo. Al igual que este, no desaparece cuando los sujetos experimentales son informados de su existencia (Wilson *et al.*, 1996), ni tampoco cuando se les ofrece dinero si logran liberarse de su influencia (Simmons *et al.*, 2010). Dado que el sesgo aparece en el derecho especialmente en el momento de establecer la condena, un posible remedio es evitar que la fiscalía haga una recomendación sobre la duración de la pena. Algo similar se podría establecer en el caso de la compensación monetaria a una víctima en un juicio de responsabilidad. Si el demandante no tiene la posibilidad de exigir una cantidad determinada de dinero, el juez tendrá la libertad de determinar esa cuantía sin correr el riesgo de obedecer al ancla sentada por el demandante. No es suficiente imponer unos topes a ciertos tipos de demandas, porque los topes se convierten en anclas inmediatamente. Las soluciones planteadas para las compensaciones y las penas, sin embargo, implican una reconcepción radical de la normatividad existente en muchas jurisdicciones.

3. La racionalidad acotada

Las consideraciones anteriores hacen evidente que las decisiones judiciales están afectadas de principio a fin por sesgos cognitivos, y que la posibilidad de eliminarlos por completo es nula. ¿Qué implicaciones tiene esto para la legitimidad racional de esas decisiones? Para poder responder a esta

pregunta es necesario aclarar el significado de la expresión "legitimidad racional".

En primer lugar, la legitimidad racional que tengo en mente no se desprende de las reglas de racionalidad para la argumentación jurídica (Atienza, 1991), pues cualquier conjunto de reglas racionales de argumentación se puede cumplir a cabalidad sin solucionar ninguno de los problemas derivados de la influencia de los sesgos cognitivos. Consideremos, por ejemplo, el catálogo de reglas de racionalidad argumentativa presentada por Bernal (2012: 107): (1) claridad y consistencia conceptual, (2) consistencia normativa, (3) saturación, (4) obediencia a la lógica deductiva, (5) obediencia a las cargas de la argumentación, (6) honestidad, (7) consistencia argumentativa y (8) coherencia. Si las premisas fácticas de un razonamiento que justifica una decisión judicial han sido afectadas por un sesgo cognitivo, el cumplimiento de estas ocho reglas de racionalidad no va a impedir que la decisión final sufra una deformación en el proceso.

Los principios de racionalidad de los que se desprende la legitimidad de las decisiones judiciales deben tener en cuenta las limitaciones psicológicas que hemos venido discutiendo en secciones anteriores. Durante las últimas décadas, en gran medida como crítica a la teoría de la elección racional utilizada en economía, ha surgido la idea de buscar un concepto de racionalidad más consecuente con las limitaciones del razonamiento humano. La idea de una racionalidad limitada o acotada (*bounded rationality*), propuesta por Herbert Simon (1956), ha servido como base para una reformulación de las exigencias que se le deben imponer a cualquier proceso de decisión. La racionalidad acotada se entiende de varias maneras en la literatura contemporánea[5], pero a grandes rasgos es una visión de la racionalidad moldeada por los resultados empíricos de la psicología. Más

---

5 Existe incluso una disputa acerca de quiénes son los verdaderos herederos del proyecto original de Simon. Gigerenzer y sus colegas (Gigerenzer y Selten, 2001) sostienen que las heurísticas "rápidas y frugales" que ellos defienden como reemplazo de las normas de racionalidad de la teoría racional de la decisión representan la verdadera continuación de la idea original, y rechazan el uso del término "racionalidad acotada" como sinónimo de la idea de "optimización bajo restricciones" utilizada en economía (Sargent, 1993) o de la racionalidad estudiada dentro de la tradición de sesgos y heurísticas de Kahneman y Tversky que he venido discutiendo en este ensayo. La

precisamente, es una forma de estudiar empíricamente cómo nuestros juicios y decisiones se apartan sistemáticamente de ideales normativos establecidos a priori, examinando tanto los procesos psicológicos subyacentes como las influencias del ambiente. Sin embargo, no es del todo claro en la literatura cómo se puede inferir un resultado *prescriptivo* a partir de los resultados *descriptivos* de la psicología (Persad, 2014). En otras palabras, ¿qué es lo que nos permite decir que actuar de la manera en que lo revela la psicología experimental es una equivocación? La respuesta, por lo menos en el caso de los agentes judiciales, es que cualquier decisión que revele un sesgo es una violación del ideal de imparcialidad que rige la administración de justicia. Los sesgos generan un tipo de parcialidad involuntaria, y la psicología nos dice cuál es la gravedad de esa parcialidad y las circunstancias que la aumentan o aminoran.

La racionalidad acotada no debe ser confundida con un concepto similar propuesto por Alexy (1989). El discurso jurídico, según Alexy, es una subcategoría o caso especial del discurso práctico en general. Además de las reglas generales que garantizan la racionalidad del discurso práctico, el discurso jurídico está restringido por condiciones operativas que se derivan de la doctrina jurídica originaria prevalente y de las reglas procesales válidas. Además, en las diferentes ramas del derecho se añaden elementos como las reglas procesales específicas, la distribución de la carga de la prueba, los tiempos procesales y las asimetrías probatorias entre las partes en los procesos penales. Las reglas de racionalidad del discurso jurídico sirven de base para darles una *justificación acotada* a las decisiones jurídicas: "Las decisiones prácticas ordinarias pretenden ser racionalmente justificables. Las decisiones jurídicas, sin embargo, dan pie a una pretensión más limitada, la de ser racionalmente justificables dentro del marco de un orden jurídico válido" (Aarnio, Alexy y Peczenik, 1981: 275; trad. propia).

Ahora bien, como los sesgos cognitivos afectan el razonamiento práctico en todas las esferas de la vida humana, la justificación racional en el derecho estaría doblemente acotada, tanto por el contexto normativo del

---

descripción que proporciono en el texto intenta ser lo más neutral posible con respecto a esta disputa terminológica.

derecho como por las limitaciones del juez en términos de memoria y capacidad cognitiva. Las reglas de racionalidad del discurso práctico en general propuestas por Alexy no garantizan la eliminación de este segundo tipo de limitaciones a la justificación racional.

Mi propuesta es que invirtamos el análisis de Alexy, y en lugar de ver a la doctrina jurídica y a las reglas procesales como limitaciones adicionales a la racionalidad práctica, encontremos en ellas más bien una fuente adicional de justificación racional que subsane las falencias de la racionalidad práctica en general. Puesto de otro modo, si los sesgos cognitivos son un elemento deslegitimador en la toma de decisiones en general, las reglas procesales deben intentar subsanar las limitaciones del razonamiento práctico imponiendo condiciones que restrinjan el efecto negativo de esos sesgos. La decisión a la que se llegue no será óptima en un sentido absoluto de racionalidad, pero será satisfactoria dentro de las limitaciones cognitivas de los jueces. Las decisiones judiciales exitosas serán producto del trabajo conjunto de mentes con limitaciones cognitivas que se adaptan a un ambiente normativo a través de reglas procesales que producen resultados satisfactorios. La tarea que se debe emprender, entonces, es la de buscar esas reglas procesales que disminuyan el riesgo de que los sesgos deformen la valoración de las pruebas o la toma de decisiones.

4. LA ELIMINACIÓN DE SESGOS A TRAVÉS DEL DERECHO

La idea de utilizar las reformas a las normas como una manera de prevenir los errores generados por los sesgos cognitivos en la vida diaria ha sido desarrollada durante los últimos años combinando las lecciones de la economía conductual y el derecho. Las propuestas más conocidas son las de Sunstein, Jolls y Rachlinski[6]. El proyecto, al cual se refieren como "*debiasing through law*", que puede ser traducido como "eliminación de sesgos a través del derecho", puede ser desarrollado de dos maneras diferentes. La primera es imponer medidas restrictivas para limitar el daño generado por la racionalidad acotada. Por ejemplo, existe bastante evidencia

---

6 Véase SUNSTEIN, JOLLS y THALER, 1998; KOROBKIN y ULEN, 2000; SUNSTEIN, 2000; RACHLINSKI, 2000; TROUT, 2005; JOLLS y SUNSTEIN, 2006a, 2006b.

de que la gente tiende a pensar que los productos o comportamientos que representan un riesgo para su salud son menos dañinos de lo que en realidad son. En esos casos el Estado y las agencias reguladoras pueden intervenir para limitar el daño a través de medidas preventivas. El ejemplo más famoso es la restricción impuesta por el alcalde Bloomberg en Nueva York en 2012 al tamaño máximo en que pueden ser vendidas las bebidas azucaradas en los locales de comida rápida[7].

La otra aproximación al problema es permitir que la gente ejerza su autonomía y sus libertades individuales, pero moldeando el ambiente para encauzar sus escogencias en ciertas direcciones, dándole un *nudge* o pequeño empujón (THALER y SUNSTEIN, 2008). Así, en lugar de restringirle al consumidor el tamaño de las bebidas gaseosas que pueda pedir, es posible alterar lo que Thaler y Sunstein llaman la "arquitectura de opciones" (p. 3) de tal forma que las personas tiendan a escoger las bebidas más saludables o más pequeñas sin que sean obligadas a hacerlo. Las decisiones de las personas serán mejores, no porque reformemos a las personas, sino porque reformamos el contexto de aprendizaje y la forma en que son presentadas las opciones. Por ejemplo, hasta hace no muchos años era normal incluir una fotografía en la hoja de vida al solicitar un trabajo. Las imágenes son una de las fuentes más conocidas de sesgos en estos contextos debido al efecto halo y a los prejuicios raciales. Una medida que evita la aparición de sesgos es la venta de formularios estándar de hoja de vida que no tengan un espacio para pegar una fotografía. La persona podrá anexar una fotografía si así lo desea, pero la tendencia será a no incluir una fotografía. Los ejemplos son numerosos y el uso de *nudges* jurídicos en diversas áreas de la actividad humana es un fructífero campo de estudio en la actualidad.

Ahora bien, las técnicas de eliminación de sesgos a través del derecho se han dirigido en su gran mayoría a cambiar el contexto normativo para mejorar las decisiones del ciudadano del común. Cuando se han aplicado al derecho mismo, se han concentrado en reformas que eliminen los sesgos

---

7 Bloomberg modificó el artículo 81 del Código de Salubridad de la Ciudad de Nueva York para que el tamaño máximo fuera 475 ml. Tras una batalla en las cortes, la medida fue finalmente rechazada por la Corte de Apelaciones del Estado de Nueva York en 2015.

en los jurados (SAKS y KIDD, 1980). Por ejemplo, la regla 407 de las FRE mencionada en la sección anterior tiene este propósito, al igual que la regla 404, que no permite que se utilice en un juicio información sobre condenas anteriores debido al peligro de generar un prejuicio desmedido contra el acusado. Por otra parte, se han modificado reglas procesales que impedían llegar más fácilmente a conciliaciones en casos civiles debido al sesgo de exceso de confianza (BABCOCK *et al.*, 1997). Sin embargo, existen muy pocos estudios acerca de cómo utilizar esta aproximación para mejorar las decisiones de los jueces. Esto se puede explicar, paradójicamente, a través de un sesgo cognitivo: el sesgo del punto ciego (PRONIN *et al.*, 2002). Según este sesgo, los seres humanos tendemos a ver en los demás sesgos que somos incapaces de ver en nosotros mismos. Los jueces que han emprendido la tarea de reformar los códigos para disminuir las decisiones sesgadas de los jurados no han emprendido una tarea semejante para disminuir sus propios sesgos. El efecto del sesgo del punto ciego es tan pronunciado que cuando se trata de casos sin jurados muchos jueces federales en Estados Unidos prescinden por completo de las *Federal Rules of Evidence*:

> Debido a que la mayoría de las reglas de evidencia son esencialmente exclusionarias, y como los jueces (como el resto de nosotros) no quiere aislarse de información potencialmente útil, muchos jueces tratan al derecho probatorio como un estorbo contraproducente que debe ser evitado cuando sea posible (SCHAUER, 2006: 1090)[8].

Sobra decir que tal actitud exacerba el problema de la deslegitimación racional de sus decisiones al abrirle la puerta completamente a la influencia de los sesgos cognitivos.

La pregunta que quiero considerar en el resto del capítulo es: ¿en qué consistiría la eliminación de sesgos a través del derecho aplicada a los jueces? La propuesta debe cumplir dos condiciones. En primer lugar,

---

8 El sustento teórico de esta actitud es, por supuesto, la obra de Bentham y su defensa de la libertad probatoria (*free proof*). El presente ensayo puede ser visto como un ataque a la conclusión de Bentham en *A Treatise on Judicial Evidence*: "*no excluyas ninguna prueba o testimonio sólo por el temor a ser engañado*" (1825: 255; cursiva original, trad. propia).

debe estar basada en evidencia empírica que muestre cuáles son las estrategias o métodos que efectivamente reducen los errores generados por los sesgos. Es necesario advertir que el estudio de las técnicas de eliminación de sesgos todavía está en una etapa naciente, pero esto no debe ser visto del todo como un defecto puesto que el derecho puede ser uno de sus terrenos de prueba más fructíferos. En segundo lugar, debe ir mucho más allá de simples recomendaciones o advertencias a los jueces. Como vimos anteriormente, el simple conocimiento de la posible influencia de un sesgo tiene un efecto nulo sobre las personas. Esto implica que la intervención debe ser directamente sobre la normatividad y el diseño institucional. Es importante tener en cuenta que es imposible eliminar completamente cualquiera de estos sesgos, como hemos visto, y que por lo tanto el propósito debe limitarse a disminuir su efecto. A continuación examinaré y rechazaré dos posibles soluciones al problema, y propondré en términos generales cuál creo que debe ser la dirección en la que se ha de buscar una reforma a las reglas procesales.

Una primera posibilidad para atenuar el efecto de los sesgos cognitivos es la introducción, para algunas decisiones, de reglas de predicción estadística (RPS). BISHOP y TROUT (2005) han mostrado que, partiendo de la misma evidencia, "las predicciones de las RPS son al menos tan confiables, y típicamente más confiables, que las predicciones de un experto humano" (p. 12). Entre muchos otros usos, las RPS han sido utilizadas para predecir la probabilidad de reincidencia de un prisionero en las audiencias para decidir su libertad condicional. A pesar de que algunas RPS aparentemente han sido exitosas en este campo, la gran mayoría de estas herramientas sufren de problemas metodológicos que hacen que su utilidad sea muy cuestionable (PÁEZ, 2016). Por ejemplo, un estudio reciente (DRESSEL y FARID, 2018) mostró que COMPAS, un software comercial de predicción de riesgo de reincidencia[9], no es más preciso que las predicciones hechas por personas con poca o ninguna experiencia criminológica. Además, a pesar de que COMPAS utiliza 137 parámetros, la misma precisión se puede

9 COMPAS es el acrónimo de "*Correctional Offender Management Profiling for Alternative Sanctions*". Fue introducido en 1998 y ha sido utilizado para decidir la libertad condicional de más de un millón de personas en Estados Unidos y otros países.

lograr con un simple modelo linear de solo dos parámetros. Si una RPS falla en una tarea tan básica, es muy dudoso que sea posible utilizar métodos semejantes en decisiones más complejas dentro de un juicio.

Una segunda alternativa para lidiar con las limitaciones cognitivas de los jueces es el uso de heurísticas "rápidas y frugales". En lugar de buscar mejores reglas procesales, el movimiento liderado por Gigerenzer (GIGERENZER y ENGEL, 2006) busca encontrar heurísticas simples dentro del derecho que permitan llegar fácilmente a decisiones correctas[10]. A primera vista esto parece un contrasentido, teniendo en cuenta que gran parte de los errores y sesgos que ocurren al juzgar y decidir son el resultado de las heurísticas que utilizamos. Además, la simplicidad del tipo de heurísticas que busca Gigerenzer parece contraria a la idea muy extendida en el derecho de que siempre se debe hacer una valoración de todos los medios probatorios relevantes. Sin embargo, la idea de Gigerenzer es encontrar empíricamente cuáles heurísticas tienen una utilidad demostrada.

Consideremos un ejemplo de una heurística rápida y frugal. Según la heurística conocida como *Take The Best* (escoge el mejor), las personas están buscando constantemente señales (*cues*) que les ayuden a tomar decisiones. Verbigracia, si comparamos dos ciudades y queremos saber cuál tiene mayor población, una de las señales que se puede utilizar es el hecho de que una de las ciudades tiene un equipo de fútbol en una liga profesional. En Alemania, por ejemplo, si comparamos una ciudad con equipo de fútbol en la *Bundesliga* y otra sin él, en el 87% de los casos la que tiene equipo de fútbol tiene mayor población. La *validez* de una señal es la frecuencia con la que esta corresponde al criterio buscado. En este caso la validez de la señal del equipo de fútbol tiene una validez de 0,87 para el criterio de mayor población. La heurística *Take The Best* funciona de la siguiente manera: al tomar una decisión, las personas usan primero la señal con mayor validez; si el criterio es inaplicable, siguen con la siguiente señal en orden descendente hasta obtener una respuesta. La información del resto de las señales es desechada (GIGERENZER y GOLDSTEIN, 1999).

---

10 LERNER (2006) defiende el uso de intuiciones y corazonadas en el trabajo policial con base en las ideas de Gigerenzer.

¿Cómo se podría utilizar una heurística como *Take The Best* en un juicio? Un juez podría utilizar cualquiera de las siguientes señales: el testimonio de un testigo identificando al culpable, la confesión del acusado, el testimonio de un policía o el testimonio de un perito. A juzgar por el comportamiento de jueces y jurados, no hay una señal más válida para condenar a una persona que la identificación por parte de un testigo. La confesión del acusado ocupa un lugar ligeramente inferior en la escala de validez, mientras que los testimonios del policía y el perito estarían un poco más abajo. Ahora bien, para que una heurística como *Take The Best* funcione, la validez de las señales debe tener una base empírica adecuada: "No podemos esperar que una heurística inferencial haga inferencias correctas (verdaderas) si su *input* o sus premisas son sustancialmente erróneas. 'Basura adentro, basura afuera', como dice el refrán" (GOLDMAN, 2003: 219). En efecto, varios estudios indican que la validez subjetiva que los jurados les asignan a estas señales no coincide con su validez objetiva. La identificación por parte de un testigo, por ejemplo, está sometida a varios sesgos de la memoria que hacen que su confiabilidad sea más baja de lo que generalmente se piensa (LOFTUS y DOYLE, 1997; BENFORADO, 2015). La situación es peor aún con las confesiones. A raíz del bien conocido *Innocence Project*, se han descubierto muchas confesiones forzadas que llevaron a condenas erróneas. En el caso de los jueces es razonable asumir que parte de su formación incluye estudiar la confiabilidad real de estas señales. Supongamos por un momento que un juez hace una valoración adecuada de la validez de las señales. Si la señal que ha sido mejor clasificada no supera el estándar de prueba, no tiene sentido continuar con las demás señales, como lo ordena la heurística, puesto que naturalmente las demás tampoco lo harán. Sin embargo, una valoración en conjunto de todas las pruebas puede llevar al juez a decidir que sí se ha superado el estándar de prueba. Desechar casi todo el acervo probatorio, a favor y en contra del acusado, y poner todo el peso de la decisión en una sola señal, iría en contra de uno de los principios rectores del derecho sin que se vea claramente cuál es el beneficio de hacerlo[11]. Finalmente, si en un caso no es posible utilizar

11 Es necesario advertir, sin embargo, que existe evidencia de que los decisores muchas veces se concentran solo en algunas de las pruebas, y no necesariamente en las más

ninguna de las señales que sirven de base para la inferencia heurística, el decisor debe volver a las normas procesales tradicionales, con todas las limitaciones racionales que hemos visto. En breve, el método heurístico no es confiable y en muchos casos será simplemente inaplicable[12].

La propuesta que quiero dejar apenas esbozada aquí para la eliminación de sesgos a través del derecho está basada en la solución que se ha adoptado en medicina ante un problema semejante. Con el fin de estandarizar y mejorar la calidad de las decisiones médicas, desde hace muchos años se han venido desarrollando *guías de práctica clínica*, que identifican, resumen y evalúan toda la evidencia disponible acerca de la prevención, diagnóstico y pronóstico de un gran número de condiciones médicas, al igual que las razones en términos de costo/beneficio que subyacen a las recomendaciones para el tratamiento clínico. Estas guías sirven de base para la elaboración de *protocolos clínicos*, *algoritmos de decisión* y *listas de verificación* (*checklists*) para el tratamiento de un gran número de condiciones dentro de la práctica clínica[13].

A menudo los médicos recomiendan tratamientos que se apartan de los protocolos para una enfermedad. El argumento aducido es que cada paciente es único y que las guías de tratamiento no se aplican a todos los pacientes por igual. La verdadera razón tiene más que ver con la experiencia individual de cada médico que con la unicidad de los pacientes. El sesgo de representatividad hace que los médicos tiendan a asociar los síntomas

---

relevantes. El análisis incompleto e idiosincrásico de la evidencia también está determinado por sesgos cognitivos (Moore, 1989).

12 Esto no significa que el proyecto de Gigerenzer carezca de utilidad para el derecho. Una de sus áreas de investigación es mejorar la manera en que se presenta la información a un decisor para que la entienda mejor y cometa menos errores. Por ejemplo, el resultado de un examen médico, que bien puede ser una prueba en un juicio, es mejor entendido si es presentado en términos de frecuencias y no de probabilidades (Gigerenzer, 1996). También existen otras heurísticas menos simples y frugales que pueden ser de utilidad para el análisis de las pruebas, como veremos más adelante.

13 Algunos de estos protocolos usan gráficas y diagramas de flujo de fácil lectura. Un ejemplo muy claro es la gráfica utilizada para el tratamiento de la ictericia en recién nacidos diseñada por el National Collaborating Centre for Women's and Children's Health (2010).

de un paciente individual con los de los pacientes que han tratado exitosamente en su práctica profesional utilizando un tratamiento "alternativo" al recomendado. Desestiman el hecho de que las guías de práctica clínica están basadas en evidencia estadística que recomienda un tratamiento que tenderá a ser mucho más exitoso que el tratamiento "alternativo". Por otra parte, si el tratamiento recomendado en la guía no ha sido exitoso en los pacientes que el médico ha tratado más recientemente, el sesgo de disponibilidad lo llevará a abandonarlo en favor de otros tratamientos cuya probabilidad de éxito es menor (KOROBKIN y ULEN, 2000). A esto se debe sumar el sesgo de exceso de confianza que ataca particularmente a los médicos (GILBERT *et al.*, 1977). La existencia de las guías y de los protocolos asociados a ellas hace que los errores que se produzcan sean producto de decisiones conscientes y voluntarias, y que los profesionales de la salud puedan ser responsabilizados directamente cuando se apartan de ellos. Al mismo tiempo, seguir los protocolos clínicos tiene el efecto benéfico de proteger a los médicos frente a cualquier demanda que se derive de un resultado negativo en el tratamiento.

Mi propuesta para el caso del derecho es la creación de *guías de práctica judicial*, análogas a las guías de práctica clínica, para cada tipo de prueba y procedimiento que se lleve a cabo en el transcurso de un proceso judicial, desde la apertura de la investigación hasta la imposición de la pena. Una guía de práctica judicial tendría el propósito de identificar, resumir y evaluar toda la evidencia psicológica disponible acerca de la aparición y prevención de sesgos asociados con cada tipo de prueba, al igual que las razones en términos de costo/beneficio que subyacen a las recomendaciones para el tratamiento procesal. A partir de las guías de práctica jurídica se pueden desarrollar protocolos jurídicos, algoritmos de decisión y listas de verificación que deben ser seguidos sin excepción.

Para algunos procedimientos judiciales ya se han adoptado medidas semejantes a las que estoy proponiendo. Tomemos como ejemplo el procedimiento de reconocimiento visual de un sospechoso por parte de un testigo, ya sea a través de fotografías o en vivo. Hay muchos estudios que muestran la falibilidad de estos procedimientos de identificación. Cuando el autor del delito está en la alineación junto con otras personas inocentes, en una tercera parte de los casos el testigo no identifica a nadie; y cuando lo hace, una tercera parte de las veces identifica a una persona inocente.

Peor aún, cuando el autor del delito *no* está en la alineación, la mitad de las veces el testigo escoge a alguno de los integrantes inocentes de la alineación en lugar de decir que no identifica a ninguno. En Estados Unidos, de las primeras 250 condenas erróneas que fueron reversadas utilizando pruebas de ADN, 190 de ellas se debieron a la identificación errónea del condenado por parte de un testigo (BENFORADO, 2015: 112-113).

En una guía de práctica judicial sobre el reconocimiento visual este diagnóstico del problema debe ir acompañado de una explicación teórica que dé cuenta de él. La causa reside, por supuesto, en la fragilidad y los sesgos de nuestra memoria. No viene al caso presentar aquí la evidencia psicológica que explica estos errores cognitivos, y cómo estos son exacerbados por elementos contextuales[14], pero debería ser tarea del ministerio público procesar este enorme caudal de información para poder extraer protocolos para reducir el riesgo de error. Por ejemplo, muchas veces los oficiales de policía intencional o desprevenidamente le dan señales al testigo para identificar al sospechoso. La solución es que el policía que administra la alineación no sepa quién es el sospechoso. En ocasiones las personas inocentes que son incluidas en la alineación no tienen ninguna semejanza con el sospechoso. La solución obvia es asegurarse de que dicha similitud exista. Además, los testigos tienden a sentirse presionados para identificar a alguien y pueden terminar identificando al individuo más parecido al que recuerdan, así no estén completamente seguros. La solución es advertir al testigo que el sospechoso puede no estar en la alineación que va a presenciar y que la investigación continuará, independientemente del resultado de esa diligencia. Finalmente, está comprobado que después de siete días hay una caída dramática en la fidelidad de nuestros recuerdos. La solución es lograr que el proceso de reconocimiento se lleve a cabo durante la primera semana.

Algunas jurisdicciones han tomado medidas que incorporan estas soluciones. El estado de Nueva Jersey introdujo en 2001 un estricto protocolo para el reconocimiento visual (State of New Jersey, 2001). En Chile se creó el Protocolo Interinstitucional de Reconocimiento de Imputados,

---

14 EPSTEIN (2014) reporta que en los últimos 30 años se han publicado más de 2.000 artículos sobre el tema.

publicado por el Ministerio Público (2013), el cual incluye muchas de las recomendaciones mencionadas en el párrafo anterior[15]. El uso de este tipo de protocolos tiene el potencial de reducir el riesgo de error, siempre y cuando el diagnóstico del problema y el análisis teórico y experimental sean correctos. Todos estos son los elementos que deben constituir una guía de procedimiento judicial.

Ahora bien, el ejemplo anterior es relativamente simple y los usuarios de la guía serían oficiales de policía. La pregunta es si es posible generar guías de procedimiento judicial y protocolos para el actuar de los jueces en el contexto del juicio. Consideremos un ejemplo más complejo que involucra a los jueces, el caso de las pruebas periciales. Existe una amplia literatura en el derecho estadounidense en torno al caso *Daubert v. Merrell Dow* (1993) y la función de cancerberos del conocimiento científico que este les asigna a los jueces. Durante los 25 años en que ha servido de base para la admisibilidad del conocimiento científico a nivel federal se han documentado innumerables casos en que los jueces se apartan de los principios estipulados en la sentencia, ya sea para rechazar conocimiento científico perfectamente válido o para admitir pruebas de dudoso origen. Beecher-Monas (2000) sostiene que el uso de un protocolo puede asegurar que el juez lleve a cabo un análisis juicioso de las pruebas periciales:

> Primero, el juez debe identificar y entender la teoría y la hipótesis subyacente. Segundo, para poder determinar si la teoría es defendible, el juez debe examinar toda la información disponible […] Tercero, en donde haya vacíos de información (que son inevitables), los jueces deben llenarlos con supuestos por defecto que sean científicamente justificables. Cuarto, el juez debe investigar la metodología […] y determinar si se ajusta a las prácticas generalmente aceptadas en su campo. Finalmente, el juez debe unir toda esta información de tal forma que pueda hacer una evaluación probabilística de la fortaleza de los vínculos entre teoría, supuestos, metodología y la conclusión defendida por el perito (p. 1590; trad. propia).

Ahora bien, muchas de las recomendaciones que se desprenderán de las guías de procedimiento judicial irán encaminadas a impedir que el juez

---

15 Duce (2017) hace un análisis en profundidad del problema en el caso chileno.

tenga acceso a cierta información sesgadora. Esto se podría lograr, como lo sugiere SCHAUER (2006: 201), separando la función de admitir las pruebas de la función de decidir acerca de la culpabilidad del imputado con base en las pruebas admitidas, tal y como ocurre en los sistemas con jurados, en los que el juez decide la admisibilidad y el jurado decide la culpabilidad. Naturalmente esto requeriría de un cambio profundo en el diseño institucional, pero tendría la ventaja de preservar en alguna medida la libertad decisoria de los jueces, asegurando al mismo tiempo que las pruebas que utiliza para su decisión han sido filtradas por personas sin poder de decisión y que han seguido protocolos estrictos tendientes a disminuir el efecto negativo de los sesgos cognitivos.

Finalmente, es perfectamente posible que algunos de los protocolos que se deriven de la elaboración de guías de procedimiento judicial sean iguales o muy similares a las reglas de procedimiento utilizadas en la actualidad[16]. Al fin y al cabo, el derecho ha tenido un conocimiento intuitivo de la existencia de algunos de estos sesgos desde hace mucho tiempo. Pero, aunque en algunos casos al final del análisis lleguemos al mismo punto de partida, esto no debe ser visto como una pérdida de tiempo y esfuerzo. El ejercicio de encontrar la base empírica que justifica la adopción de una regla procesal le otorga la legitimidad racional que no le puede dar la tradición. Y si esto es cierto de las reglas procesales existentes, con mayor razón lo será de las nuevas reglas y protocolos que se desprendan directamente del estado del arte en psicología, compendiado en las guías de procedimiento judicial.

CONCLUSIONES

El derecho ha podido desestimar la importancia de los sesgos cognitivos debido a la falta de mecanismos de autoevaluación y de revisión de la calidad epistémica de sus decisiones. Un diálogo fructífero entre el derecho y la psicología puede llevar a mejorar la calidad de los sistemas judiciales a través del mejoramiento de las decisiones que se toman en cada etapa de

16 Hay que tener en cuenta, además, que no todas las reglas procesales tienen como objetivo alcanzar la verdad. En ellas se mezclan objetivos encaminados a la preservación de los derechos de las partes y en algunos casos no es factible separar las dos dimensiones.

los procesos. La simple exigencia de que los jueces expongan las razones de sus decisiones no será suficiente para lograr ese cometido, ni tampoco el uso de heurísticas simples que pueden tener el efecto contrario y aumentar el número de errores. La sugerencia de este capítulo es que la única forma de contrarrestar los efectos negativos de los sesgos cognitivos en los jueces es a través de la creación de guías de práctica judicial que puedan servir de base para la creación de protocolos procesales que limiten en la medida de lo razonable sus efectos negativos. Esta es una tarea que requiere de un conocimiento profundo del derecho procesal y de la psicología cognitiva. Es una combinación difícil de encontrar, pero que se debe estimular en aras de mejorar la racionalidad de las decisiones judiciales[17].

El tipo de aproximación presentado en este capítulo ha sido catalogado como "paternalista" por algunos autores (Gigerenzer y Engel, 2006: 9). Trout afirma justamente lo contrario:

> Eliminar los sesgos [a través del derecho] promueve, no socava, la autonomía. Después de todo, los sesgos amenazan nuestra habilidad de alcanzar nuestras metas sopesadas de largo plazo. Al reconocer la amenaza podemos aumentar la probabilidad de alcanzar nuestras metas de largo plazo (Trout, 2005: 414; trad. propia).

Por otra parte, una propuesta que sugiere la introducción de protocolos y reglas procesales que restringen la autonomía del juez encontrará resistencia en la práctica por parte de juristas de espíritu benthamiano. Al fin y al cabo, son medidas que atacan en cierta medida la "libre valoración" de las pruebas de acuerdo con las reglas de la "sana crítica". Pero si la valoración no es tan "libre" y la crítica no es tan "sana" debido a la presencia recurrente de sesgos cognitivos, sería irracional interponer obstáculos doctrinales si se quiere alcanzar un mayor grado de legitimidad racional para las decisiones judiciales.

---

17 Para una defensa de esta aproximación naturalista al derecho probatorio, véase Leiter y Allen, 2001.

REFERENCIAS

Aarnio, A., Alexy, R. y Peczenik, A., 1981: "The foundation of legal reasoning", *Rechtstheorie*, 21: 133-158, 257-279, 423-448.

Alexy, R., 1989: *A theory of Legal Argumentation: The Theory of Rational Discourse as Theory of Legal Argumentation*. Oxford: Clarendon Press.

Amaya, A., 2013: "Coherence, evidence, and legal proof", *Legal Theory*, 19: 1-43.

Amaya, A., 2015: "Coherencia, virtud y prueba en el derecho", en Páez, A. (ed.), *Hechos, evidencia y estándares de prueba*. Bogotá: Ediciones Uniandes, 37-63.

Atienza, M., 1991: *Las razones del derecho. Teorías de la argumentación jurídica*. Madrid: Centro de Estudios Constitucionales.

Ask, K. y Pina, A., 2011: "On being angry and punitive: how anger alters perception of criminal intent", *Social Psychological and Personality Science*, 2(5): 494-499.

Babcock, L., Loewenstein, G. e Issacharoff, S., 1997: "Creating convergence: debiasing biased litigants", *Law and Social Inquiry*, 22: 913-925.

Beecher-Monas, E., 2000: "The heuristics of intellectual due process: a primer for triers of science", *NYU Law Review*, 75: 1563-1657.

Benforado, A., 2015: *Unfair. The New Science of Criminal Injustice*. New York: Crown Publishers.

Bentham, J., 1825: *A Treatise on Judicial Evidence*, extracted from the manuscripts of Jeremy Bentham by M. Dumont. London: J. W. Paget.

Bernal, C., 2012: "Legal argumentation and the normativity of legal norms", en Dahlman, C. y Feteris, E. (eds.), *Legal Argumentation Theory: Cross-Disciplinary Perspectives*. Dordrecht: Springer, 103-112.

Bishop, M. y Trout, J. D., 2005: *Epistemology and the Psychology of Human Judgment*. New York: Oxford University Press.

CONLISK, J., 1996: "Why bounded rationality?", *Journal of Economic Literature*, 34: 669-700.

Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).

DE LA ROSA RODRÍGUEZ, P. y SANDOVAL, V. D., 2016: "Los sesgos cognitivos y su influjo en la decisión judicial. Aportes de la psicología jurídica a los procesos penales de corte acusatorio", *Derecho Penal y Criminología*, 38: 141-164.

DI LELLO, E. V., 1993: "Fighting fire with firefighters: a proposal for expert judges at the trial level", *Columbia Law Review*, 93: 473-507.

DRESSEL, J. y FARID, H., 2018: "The accuracy, fairness, and limits of predicting recidivism", *Science Advances*, 4, eaao5580.

DUCE, M., 2017: "Reconocimientos oculares: una aproximación empírica a su funcionamiento y algunas recomendaciones para su mejora", *Política Criminal*, 12: 291-379.

EPSTEIN, J., 2014: "Eyewitness and erroneous convictions: an American conundrum", en COOPER, S. L. (ed.), *Controversies in Innocence Cases in America*. Burlington: Ashgate.

FARINA, F., ARCE, R. y NOVO, M., 2002: "Heurístico de anclaje en las decisiones judiciales", *Psicothema*, 14: 39-46.

FISCHHOFF, B., SLOVIC, P. y LICHTENSTEIN, S., 1978: "Fault trees: sensitivity of estimated failure probabilities to problem representation", *Journal of Experimental Psychology: Human Perception and Performance*, 4: 330-344.

FRANK, J., 1949: *Law and the Modern Mind*. London: Stevens y Sons.

GILBERT, J. P., MCPEEK, B. y MOSTELLER, F., 1977: "Statistics and ethics in surgery and anesthesia", *Science*, 198: 684-689.

GIGERENZER, G., 1996: "The psychology of good judgment: frequency formats and simple algorithms", *Medical Decision Making*, 16: 273-280.

GIGERENZER, G. y GOLDSTEIN, D. G., 1999: "Betting on one good reason: the Take The Best heuristic", en GIGERENZER, G., TODD, P. M. y The ABC Research Group, *Simple Heuristics that Make Us Smart*. New York: Oxford University Press, 75-95.

GIGERENZER, G. y SELTEN, R. (eds.), 2001: *Bounded Rationality. The Adaptive Toolbox*. Cambridge: MIT Press.

GIGERENZER, G. y ENGEL, C., 2006: *Heuristics and the Law*. Cambridge: MIT Press.

GIGERENZER, G., TODD, P. M. y The ABC Research Group, 1999: *Simple Heuristics that Make Us Smart*. New York: Oxford University Press.

GOLDMAN, A. I., 2003: "Simple heuristics and legal evidence", *Law, Probability, and Risk*, 2: 215-226.

Graham v. John Deere Co. 383 U.S. 1 (1966).

GUTHRIE, C., RACHLINSKI, J. J. y WISTRICH, A. J., 2001: "Inside the judicial mind", *Cornell Law Review*, 86: 777-830.

HASTIE, R., SCHKADE, D. A. y PAYNE, J. W., 1999: "Juror judgments in civil cases: hindsight effects on judgments of liability for punitive damages", *Law and Human Behavior*, 22: 597-614.

In re Chamberlain's Estate, 156A. 42, 43 (N.J. Prerogative Ct. 1931).

JOIS, G. U., 2009: "Stare decisis is cognitive error", *Brooklyn Law Review*, 75: 63-141.

JOLLS, C. y SUNSTEIN, C., 2006a: "Debiasing through law", *The Journal of Legal Studies*, 35: 199-241.

JOLLS, C. y SUNSTEIN, C., 2006b: "The law of implicit bias", *California Law Review*, 94: 969-996.

KAHNEMAN, D. y TVERSKY, A., 1973: "On the psychology of prediction", *Psychological Review*, 80: 237-251.

KAHNEMAN, D., SLOVIC, P. y TVERSKY, A. (eds.), 1982: *Judgment under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

KOROBKIN, R. B. y ULEN, T. S., 2000: "Law and behavioral science: removing the rationality assumption from law and economics", *California Law Review*, 88: 1051-1144.

LABINE, S. J. y LABINE, G., 1996: "Determinations of negligence and the hindsight bias", *Law and Human Behavior*, 20: 501-516.

LAUDAN, L., 2006: *Truth, Error and Criminal Law*. New York: Cambridge University Press.

LEITER, B. y ALLEN, R. J., 2001: "Naturalized epistemology and the law of evidence", *Virginia Law Review*, 87: 1491-1550.

LEMPERT, R. O., 1977: "Modelling relevance", *Michigan Law Review*, 75: 1021-1057.

LERNER, C. S., 2006: "Reasonable suspicion and mere hunches", *Vanderbilt Law Review*, 59: 407-473.

LOFTUS, E. F. y DOYLE, J., 1997: *Eyewitness Testimony: Civil and Criminal*. Dordrecht: Kluwer.

LYON, T. D. y KOEHLER, J. J., 1996: "Relevance ratio: evaluating the probative value of expert testimony in child sexual abuse cases", *Cornell Law Review*, 82: 43-78.

Ministerio Público de Chile, 2013: *Protocolo Interinstitucional de Reconocimiento de Imputados*. Santiago de Chile.

MONAHAN, J. y WALKER, L., 1988: "Social science research in law: a new paradigm", *American Psychologist*, 43: 465-472.

MOORE, A. J., 1989: "Trial by schema. Cognitive filters in the courtroom", *UCLA Law Review*, 37: 273-341.

MUÑOZ ARANGUREN, A., 2011: "La influencia de los sesgos cognitivos en las decisiones jurisdiccionales: el factor humano. Una aproximación", *InDret. Revista para el Análisis del Derecho*, 2.

Myers, D. G. y Lamm, H., 1976: "The group polarization phenomenon", *Psychological Bulletin*, 83: 602-627.

National Collaborating Centre for Women's and Children's Health, 2010: *Neonatal jaundice. Clinical guideline.* London: Royal College of Obstetricians and Gynecologists.

Oldfather, C. M., 2012: "Judging, expertise, and the rule of law", *Washington University Law Review*, 89: 847-900.

Páez, A., 2016: "The prediction of future behavior: the empty promises of expert clinical and actuarial testimony", *Teoria Jurídica Contemporânea*, 1: 75-101.

Peer, E. y Gamliel, E., 2013: "Heuristics and biases in judicial decisions", *Court Review*, 49: 114-118.

Persad, G., 2014: "When, and how, should cognitive bias matter to law", *Law and Inequality*, 32: 31-67.

Posner, R. A., 1999: "An economic approach to the law of evidence", *Stanford Law Review*, 51: 1477-1546.

Pronin, E., Lin, D. Y. y Ross, L., 2002: "The bias blind spot: perceptions of bias in self versus others", *Personality and Social Psychology Bulletin*, 28: 369-381.

Rachlinski, J. J., 1998: "A positive psychological theory of judging in hindsight", *The University of Chicago Law Review*, 65: 571-625.

Rachlinski, J. J., 2000a: "Heuristics and biases in the court: ignorance or adaptation?", *Oregon Law Review*, 61-102.

Rachlinksi, J. J., 2000b: "New law and psychology: a reply to critics, skeptics, and cautious supporters", *Cornell Law Review*, 85: 739-766.

Saks, M. J. y Baron, C. H. (eds.), 1980: *The Use/Nonuse/Misuse of Applied Social Research in the Courts.* Cambridge: Abt Books.

Saks, M. J. y Kidd, R. F., 1980: "Human information processing and adjudication: trial by heuristics", *Law and Society Review*, 15: 123-160.

SARGENT, T. J., 1993: *Bounded Rationality in Macroeconomics*. Oxford: Oxford University Press.

SCHAUER, F., 2006: "On the supposed jury-dependence of evidence law", *University of Pennsylvania Law Review*, 155: 165-202.

SIMMONS, J. P., LEBOEUF, R. A. y NELSON, L. D., 2010: "The effect of accuracy motivation on anchoring and adjustment: do people adjust from provided anchors?", *Journal of Personality and Social Psychology*, 99: 917-932.

SIMON, H. A., 1956: "Rational choice and the structure of environments", *Psychological Review*, 63: 129-138.

SMALL, M. A., 1993: "Legal psychology and therapeutic jurisprudence", *St. Louis University Law Review*, 37: 675-713.

SORIA VERDE, M. A. y SÁIZ ROCA, D., 2006: *Psicología criminal*. Madrid: Pearson.

State of New Jersey, Office of the Attorney General, 2001: *Attorney General Guidelines for preparing and conducting photo and live lineup identification procedures*. April 18, 2001. Trenton, NJ.

SUNSTEIN, C. R. (ed.), 2000: *Behavioral Law and Economics*. New York: Cambridge University Press.

SUNSTEIN, C. R., 2007: "Group polarization and 12 Angry Men", *Negotiation Journal*, 23: 443-447.

SUNSTEIN, C. R., JOLLS, C. y THALER, R. H., 1998: "A behavioral approach to law and economics", *Stanford Law Review*, 50: 1471-1550.

THALER, R. H. y SUNSTEIN, C. R., 2008: *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New York: Penguin.

TROUT, J. D., 2005: "Paternalism and cognitive bias", *Law and Philosophy*, 24: 393-434.

TVERSKY, A. y KAHNEMAN, D., 1982: "Judgments of and by representativeness", en KAHNEMAN, D., SLOVIC, P. y TVERSKY, A. (eds.), *Judgment under Uncertainty: Heuristics and Biases.* New York: Cambridge University Press.

WASON, P., 1960: "On the failure to eliminate hypotheses in a conceptual task", *Quarterly Journal of Experimental Psychology*, 12: 129-140.

WEINSTEIN, I., 2003: "Don't believe everything you think: cognitive bias in legal decision making", *Clinical Law Review*, 9: 783-834.

WILSON, T. D., HOUSTON, C. E., ETLING, K. M. y BREKKE, N., 1996: "A new look at anchoring effects: basic anchoring and its antecedents", *Journal of Experimental Psychology: General*, 125: 387-402.

VII.
Un argumento naturalista en contra de la teoría retórica de la argumentación probatoria en los procesos penales

DANNY MARRERO AVENDAÑO*

INTRODUCCIÓN

La teoría retórica defiende que la finalidad de la argumentación fáctica en los procesos penales es persuadir al juez sobre la culpabilidad del acusado. Sin embargo, esta teoría no satisface nuestras expectativas sociales. Queremos que los jueces conozcan los hechos en litigio, pero la persuasión no necesariamente lleva al conocimiento. Para corroborar esta hipótesis usaré estudios de psicología aplicada que muestran que cuando las personas esperan ser evaluadas por su habilidad de construir un caso persuasivo en contra de un sospechoso son más vulnerables a los sesgos confirmatorios que cuando esperan ser evaluadas por otros criterios (como el uso de un procedimiento correcto o llegar a un resultado acertado) o ningún criterio en absoluto. En la literatura relevante, el sesgo confirmatorio es la universal, irresistible e incorregible tendencia de los seres humanos a favorecer la evidencia que justifica hipótesis, creencias y expectativas previamente adoptadas, y a ignorar, o no darle suficiente importancia, a la evidencia que las desvirtúa (NICHERSON, 1998: 175). En los fiscales e investigadores criminales, es la tendencia a concentrarse en un sospechoso produciendo, seleccionando e interpretando evidencia que podría conducir a que se le imponga una sentencia condenatoria, y pasando por alto, minimizando o desconfiando de la evidencia que lleva a una sentencia absolutoria. La principal amenaza del sesgo confirmatorio para el proceso penal es que las sentencias condenatorias erróneas no solo son atribuibles a los jueces sino también a los fiscales e investigadores criminales, que presentan casos que terminan siendo coherentes, pero que no reconstruyen apropiadamente los hechos en litigio y, en consecuencia, no orientan correctamente al adjudicador con respecto a la verdad de los hechos en contienda. En otras palabras, lo que se espera del proceso penal es que se alcance justicia a través de una investigación imparcial en la que la evidencia inculpatoria

---

* Graduado en Filosofía y en Derecho; magíster en Derecho y en Filosofía; doctorado en Filosofía de la Universidad de Arkansas (Fayetteville, Estados Unidos). Actualmente es profesor del doctorado en Ciencias Jurídicas de la Pontificia Universidad Javeriana (Bogotá, Colombia) y Course Director, Critical Thinking en Verto Education (Estados Unidos).

y exoneratoria sea evaluada objetivamente. Sin embargo, dado el sesgo confirmatorio, la evidencia exculpatoria tiende a ser minimizada, ignorada o vista como poco confiable.

Si esto es correcto, la teoría retórica debería ser reemplazada por una teoría que reduzca los sesgos confirmatorios. El objetivo principal de este escrito es sentar las bases para este reemplazo. Mi exposición seguirá el método propuesto por Alvin GOLDMAN (2003) para evaluar inferencias en procesos de adjudicación. Los pasos de dicho método son:

> Paso 1: Seleccione el proceso de inferencia objeto del análisis.
>
> Paso 2: Determine el fin, o grupo de fines, del proceso de adjudicación.
>
> Paso 3: Especifique qué tan bien el proceso de inferencia evaluado promueve el fin, o grupo de fines, del proceso de adjudicación.
>
> Paso 4: Si el proceso de inferencia en escrutinio no es efectivo o es deficiente promoviendo el fin esperado, identifique algunas alternativas para mejorarlo (ibíd.: 215).

Así, el proceso de inferencia objeto de este análisis (Paso 1) es la teoría retórica de la argumentación, en los términos expuestos en la sección 1 de este ensayo. Desde mi perspectiva, el fin del proceso de adjudicación es el de asegurar un trato sustantivamente justo a los individuos (Paso 2). Como lo defenderé en la sección 2 de este escrito, dicho objetivo se alcanza a través de la correcta aplicación de las fuentes del derecho válido y el conocimiento de los hechos en litigio. En la sección 3 mostraré cómo la teoría retórica no promueve dicho fin porque con ella no necesariamente se alcanza el conocimiento (Paso 3). Finalmente, sugeriré que se reemplace a la retórica por una teoría epistemológica de la argumentación (Paso 4). Dicha teoría defiende que el fin de la argumentación es llevar al determinador de los hechos al conocimiento (MARRERO, 2016). Como consecuencia, los fiscales y los investigadores criminales deberían ser evaluados por el estatus justificatorio de sus argumentos y no por sus habilidades persuasivas.

1. El proceso de inferencia objeto de análisis: la teoría retórica de la argumentación

Como la interpreto, la teoría retórica defiende que el fin de la argumentación en los procesos penales es persuadir al adjudicador de que el acusado es culpable del delito que se le atribuye. Mi punto de vista es corroborado por el más importante exponente de esta teoría, Chaïm Perelman (1980), quien afirma:

> El razonamiento jurídico es un caso muy elaborado de razonamiento práctico, no es una demostración formal, sino una argumentación encaminada a persuadir y convencer a sus adherentes de que una opción, decisión o actitud es preferible que otras opciones, decisiones y actitudes rivales (ibíd.: 129).

En este ensayo presentaré algunos de los postulados de la argumentación probatoria, que defiende esta teoría, a través de las ideas de una de sus defensoras contemporáneas: la teórica de la comunicación Janice E. Schuetz. Ella ha sido muy productiva en la aplicación de los principios de la teoría retórica en el análisis de casos jurídicos de alto impacto como el de Sacco y Vanzetti o el de O. J. Simpson (Schuetz, 1994, 2007).

Para empezar, ella define la argumentación fáctica en el derecho como "un tipo de razonamiento en el que se realizan inferencias a partir de la evidencia en formas que son apropiadas para el derecho" (Schuetz, 2007: 12). Dos elementos deberían ser clarificados en esta definición, a saber, "el tipo de razonamiento" que es la argumentación fáctica y las formas en las cuales este tipo de razonamiento es "apropiado para el derecho". Aclarando el primer elemento, responderé dos preguntas: ¿cuáles son los componentes constitutivos de este tipo de razonamiento?, y ¿cuál es la relación inferencial que se establece entre estos componentes? Siendo un tipo de razonamiento, la argumentación fáctica en el derecho está constituida por "una expresión verbal que incluye una afirmación y el soporte a dicha afirmación" (véase Diagrama 1). La afirmación que es soportada se relaciona con una atribución de responsabilidad penal, y la oración u oraciones que sirven de soporte son la evidencia que se recolecta en un caso específico. Usando el ejemplo de Schuetz:

> Un abogado podría afirmar que el acusado es culpable de conducir embriagado, basándose en la evidencia de que en una muestra de sangre del acusado se encontraron rastros de alcohol, en que hay testigos que afirman cómo el acusado conducía erráticamente, y en que hay manifestaciones visibles de que el acusado tenía un comportamiento propio de alguien que está borracho, como su dificultad para hablar y su incapacidad de caminar en una línea recta (ibíd.: 12).

Identificando el vínculo inferencial entre la atribución de responsabilidad y la evidencia, Schuetz afirma que la argumentación fáctica en el derecho es un razonamiento diferente de la deducción y la inducción. Desde su perspectiva, la argumentación jurídica es un tipo de razonamiento que busca "persuadir sin usar formas lógicas explícitas como el silogismo" (ibíd.: 12). Por lo tanto, la atribución de responsabilidad y la evidencia no se unen por una inferencia deductiva. De manera similar, Schuetz afirma que la argumentación fáctica tampoco es inductiva. Justifica esta idea tomando prestada la distinción de Perelman entre razonamiento práctico y teórico. Mientras el primero "incluye juicios de valor sobre la cualidad de los actos y las personas, [el segundo] se concentra en procedimientos que especifiquen una verdad probable" (SCHUETZ, 2007: 16). La argumentación fáctica no es un tipo de razonamiento teórico porque esta tiene por finalidad "persuadir a los jueces y/o jurados" (ibíd.: 13) y no intenta establecer una probable verdad, como debería hacerse por el razonamiento inductivo. En las palabras de Schuetz:

> En teoría, la finalidad de la argumentación jurídica es descubrir la verdad probable de una acción. El trabajo de Perelman muestra cómo el razonamiento práctico y el teórico convergen en la práctica jurídica. Los siguientes casos de estudio enfatizan lo práctico sobre lo teórico e identifican cómo los abogados y los jueces transforman el razonamiento teórico en razonamiento práctico usando narrativas, dramas y juegos (ibíd.: 16).

Esto quiere decir que el fin de la argumentación fáctica se alcanza a través de narrativas. Así los afirma Schuetz: "Los abogados y testigos usan narrativas para convencer a los jueces y/o a los jurados de la probabilidad y fidelidad de sus historias con relación a los juicios penales en contienda" (ibíd.: 18). Estas historias no tienen que ser verdaderas, sino que deberían

ser persuasivas. Una vez más, en las palabras de Schuetz: "Las historias presentadas por una de las partes podrían ser falsas, pero ellas persuaden a los jueces y a los jurados porque ellas parecen más creíbles y resuenan mejor con las experiencias de los [jueces o] jurados que las de las contraparte" (ibíd.: 18). De esta forma, las historias que se cuentan en los juicios penales conectan la acusación y la evidencia a través de una historia que es consistente y completa. "Los abogados construyen sus historias a partir de los fragmentos narrativos proveídos por los testigos y enfatizan segmentos de los testimonios que favorecen su teoría del caso y el tema de su historia" (ibíd.: 20).

DIAGRAMA 1. ESTRUCTURA DE LA ARGUMENTACIÓN FÁCTICA PARA SCHUETZ

Estructura del razonamiento fáctico	**Ejemplo**	**Elemento que articula dichos componentes**
Afirmación (atribución de responsabilidad)	El acusado es culpable de conducir embriagado	
↑	↑	↓ Narrativas ↑
Soporte a dicha afirmación (evidencia)	1. En una muestra de sangre del acusado se encontraron rastros de alcohol 2. Hay testigos que afirman que "el acusado conducía erráticamente" 3. Hay manifestaciones visibles de que el acusado tiene un comportamiento propio de alguien que está borracho, como su dificultad para hablar y su incapacidad de caminar en línea recta	

Continuado con las formas en las que la argumentación fáctica es "apropiada para los procesos jurídicos", es importante resaltar que Schuetz asume una teoría del derecho como un medio de solución de conflictos sociales. Ello es claro en esta cita textual: "el derecho existe para solucionar problemas sociales y regular la conducta humana" (ibíd.: 3). Esa resolución se alcanza a través de un veredicto (ibíd.: 22). En otras palabras, en los sistemas

modernos de adjudicación se espera que los determinadores de los hechos "revisen la evidencia y disciernan la inocencia o la culpabilidad del acusado, basándose en los cargos que se le atribuyen" (ibíd.: 40). Así, la condición de posibilidad de la adjudicación es la presentación de la evidencia que constituye los hechos en litigio. Usando una línea de tiempo, primero, "en los juicios penales se presenta la evidencia para establecer los hechos del caso y luego los adjudicadores deciden si esos hechos soportan la acusación" (ibíd.: 36). ¿De qué manera la argumentación fáctica promueve la resolución de conflictos a través de veredictos? Por una parte, las narrativas proveen la información necesaria para entender el conflicto bajo análisis. Schuetz afirma: "los asuntos jurídicos que se incluyen en las narrativas de los litigantes crean marcos cognitivos para que los adjudicadores los usen interpretando la ley y decidiendo los hechos de acuerdo con su coherencia, probabilidad y credibilidad" (ibíd.: 45). Por otra parte, la argumentación fáctica permite un mejor entendimiento de los alegatos de las partes, ya que "las historias les permiten a los adjudicadores conectar la información en el tema de una historia, realizar inferencias que concuerdan con las partes de la historia, y elucidar si la historia es consistente y completa" (ibíd.: 19).

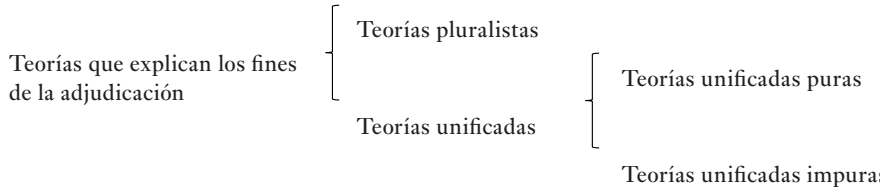
2. LOS FINES DEL PROCESO PENAL

El Paso 2 del método de análisis usado en este escrito requiere una breve clarificación. Para Goldman, las teorías que explican los fines de los procesos de adjudicación se pueden clasificar en dos grupos: teorías pluralistas y teorías unificadas (GOLDMAN, 2005: 163-164). Las teorías pluralistas defienden que los procesos de adjudicación tienen diferentes fines, ninguno de los cuales prima sobre el otro. Algunos de estos fines son la justicia, la imparcialidad, el permitir la coexistencia pacífica, la búsqueda de la verdad, la protección de los derechos. Las teorías unificadas, alternativamente, explican la adjudicación teniendo en cuenta un solo fin. Estas no defienden que los procesos de adjudicación necesariamente alcanzan ese fin, sino que lo usan como un recurso explicativo para clarificar las principales actividades adelantadas en estos. Dentro de esta segunda alternativa se pueden encontrar teorías unificadas puras y teorías unificadas impuras (véase Diagrama 2). Las teorías unificadas puras afirman que las prácticas jurídicas bajo consideración pueden ser subsumidas en un fin

específico. La teoría de Schuetz es un buen ejemplo de esta noción, ya que ella sugiere que la finalidad de los procesos de adjudicación es resolver disputas, o diferencias de opinión, y no prevé fines alternativos a ser alcanzados por el proceso jurídico. Las teorías unificadas impuras, por otra parte, defienden que aunque el fin de la adjudicación es uno, es posible identificar diferentes fines que coexisten con dicho fin principal. A modo de ejemplo, ANDERSON, SCHUM y TWINING (2005) adoptan una teoría impura unificada del procedimiento jurídico. Esto les permite defender que aunque la finalidad principal del derecho no es la determinación de la verdad, este está orientado a la búsqueda de la misma. En sus palabras:

> … el objetivo primario del derecho adjetivo (o procesal) consiste en lograr la "rectitud de la decisión" en la adjudicación, es decir, la aplicación correcta del derecho a los hechos probados como verdaderos. […] [Así] el modelo de adjudicación es, en sí mismo, instrumentalista, en tanto que la búsqueda de la verdad a través de la razón es solo un medio para lograr el fin de la justicia, considerada como la implementación del derecho sustantivo (ibíd.: 79).

DIAGRAMA 2. TEORÍAS QUE EXPLICAN LOS FINES DE LA ADJUDICACIÓN



Desde mi punto de vista, el principal problema de las teorías pluralistas es que no pueden resolver los problemas que podrían surgir entre una multiplicidad de fines sugeridos, ya que, desde esta perspectiva, todos ellos tienen la misma importancia. La falencia de las teorías unificadas puras, por su parte, es que son vulnerables a contraejemplos que muestren que los sistemas jurídicos podrían alcanzar fines diferentes a los seleccionados. A modo de ilustración, una teoría que afirma que el fin único del derecho es la solución de controversias, como la de Schuetz, sería compatible con un método de adjudicación que resuelva disputas tirando una moneda al aire, una regla que diga que quien escoja cara gana y verificando si la moneda cae del lado de la cara o del sello. Sin embargo, esto no parece

correcto porque el derecho no solamente quiere solucionar disputas, sino alcanzar justicia. Este, a su vez, es otro posible fin del sistema jurídico. Teniendo en cuenta estas críticas, creo que la mejor alternativa son las teorías impuras unificadas. Estas plantean jerarquías de fines a perseguir por el sistema jurídico, de tal forma que cuando hay conflictos entre fines jurídicos es posible darles prelación a los que tienen más importancia. Adicionalmente, estas teorías son lo suficientemente flexibles como para tener un catálogo amplio de fines subsidiarios, permitiéndole al sistema intentar alcanzar paralelamente varios ideales. Por las anteriores razones sostendré, con GOLDMAN (2005), que el fin principal de los procesos de adjudicación es asegurar un "trato sustantivamente justo a los individuos". Este fin se alcanza si el derecho válido es correctamente aplicado y si se logra el conocimiento de los hechos en litigio (ibíd.: 164).

3. ¿PROMUEVE LA TEORÍA RETÓRICA LOS FINES DE LA ADJUDICACIÓN?

Siguiendo con el Paso 3, me pregunto si la teoría retórica promueve el fin de asegurar un trato sustantivamente justo a los individuos. Para responder a esta pregunta es útil recordar que, por una parte, el conocimiento de los hechos es una condición necesaria, pero no suficiente, del fin principal de los procesos de adjudicación. También resulta útil traer a colación que la teoría retórica afirma que la finalidad de la argumentación fáctica en los procesos penales es persuadir al juez sobre la culpabilidad del acusado. Si ello es así, la teoría retórica no promueve el fin principal de la adjudicación porque ella no conduce al conocimiento. Para defender esta afirmación usaré algunos estudios sobre los efectos psicológicos de la rendición de cuentas que muestran que cuando las personas esperan ser evaluadas por sus habilidades de persuadir a otros sobre sus posiciones, se hacen susceptibles a sesgos confirmatorios. Más concretamente, usaré dos experimentos de Barbara O'Brian que "demuestran que cuando las personas esperan rendir cuentas sobre sus habilidades de persuadir a otros de la culpabilidad de un sospechoso se exacerban los sesgos cognitivos relevantes para la toma de decisiones propia de las actividades realizadas por los fiscales" (O'BRIAN, 2009a: 1). Pero antes quisiera mostrar lo ubicuo y corrosivo que es el sesgo confirmatorio en el proceso penal.

3.1. LA UBICUIDAD Y CORROSIVIDAD DEL SESGO CONFIRMATORIO EN PROCESO PENAL

Estudios empíricos muestran que los fiscales y los investigadores criminales son especialmente vulnerables a los sesgos confirmatorios (Ask y Granhag, 2005; Ask, Rabelius y Granhag, 2008; Ask, Reinhard, Marksteiner y Granhag, 2011; Burke, 2006, 2007; Eerland y Rassin, 2012; Marksterner, Ask, Reinhard y Granhag, 2011; O'Brien, 2009a; 2009 b; Rassin, Eerland y Kuijpers, 2010, entre otros). Como lo afirmé en la introducción de este ensayo, el sesgo confirmatorio es la universal, irresistible e incorregible tendencia de los seres humanos a favorecer la evidencia que justifica hipótesis, creencias y expectativas previamente adoptadas, y a ignorar, o no darle suficiente importancia, a la evidencia que las desvirtúa (Nickerson, 1998: 175). En los fiscales e investigadores criminales, es la tendencia a concentrarse en un sospechoso produciendo, seleccionando e interpretando evidencia que podría conducir a que se le imponga una sentencia condenatoria, y pasando por alto, minimizando o desconfiando de la evidencia que lleva a una sentencia absolutoria.

Permítaseme parafrasear a Keith A. Findley y Michael S. Scott (Findley y Scott, 2006) para ilustrar brevemente la forma en la que el sesgo confirmatorio permea las actividades de fiscales e investigadores criminales. Para empezar, es necesario reconocer que aunque el sesgo confirmatorio es alimentado por el procedimiento penal adversarial, este también es un *input* imperfecto del mismo. En primer lugar, el sistema adversarial polariza a sus participantes premiando a los fiscales y abogados defensores que celosamente encuentren la evidencia que los favorezca, y reconstruyan y presenten exitosamente los argumentos que justifican sus respectivas versiones de los hechos. En segundo lugar, el sesgo cognitivo es consecuencia de la información imperfecta que ingresa al sistema. Recuérdese que es en la investigación criminal en donde el sesgo confirmatorio comienza y desde donde corroe todo el sistema. La razón de ello es que la información generada en este momento procesal alimenta la totalidad del proceso. De esta forma, si lo que activa la investigación son testimonios falsos o erróneos que hacen que los investigadores criminales se centren en falsos sospechosos, el sistema tendrá un *input* defectuoso que no conducirá a la verdad, condición necesaria para la justicia requerida del sistema.

A los *inputs* imperfectos que normalmente ocurren en la etapa de investigación se suman las presiones de las víctimas, los medios de comunicación y el público en general, y las escalas de medición del trabajo de los investigadores criminales que identifican la justicia con la privación de la libertad y compelen a la policía a resolver casos rápidamente. Esta obligación compite con otras que también le son asignadas a la policía y que, en muchos casos, son incompatibles, e impiden una evaluación objetiva de la evidencia. Algunas de estas obligaciones potencialmente inconsistentes son atender apropiadamente a las víctimas de los delitos, identificar y arrestar al agresor, o salvaguardar los derechos constitucionales del sospechoso. Dicha presión también hace que se piense que la acción penal termina con la aprehensión del presunto responsable, cumpliendo un estándar de conocimiento muy bajo, como es la inferencia razonable de que ha ocurrido un delito. Sin embargo, dicho estándar no satisface los requerimientos cognitivos del juicio, que es donde se determinan los hechos y que se plantean en términos de ir más allá de la duda razonable. Finalmente, piénsese que una vez un investigador ha optado por una teoría del caso, esta condicionará el tipo de evidencia que se debería buscar y las formas en que las entrevistas e interrogatorios serán conducidos. Desde luego, ello tendrá un impacto poco trivial en la forma como el supuesto crimen es reconstruido y presentado.

Los fiscales también son presa del sesgo confirmatorio. Mientras la presión sobre la policía está en resolver casos arrestando rápidamente a los sospechosos, las evaluaciones laborales de los fiscales se centran en números de acusaciones, sospechosos privados de la libertad de forma preventiva y sentencias condenatorias. Esta presión se aúna al hecho de que los fiscales no perciben los hechos objeto de la acusación directamente, sino que los reconstruyen gracias a inferencias que realizan a través de la evidencia que ha sido recolectada y de la información parcial que les presentan los investigadores y los testigos presenciales. Finalmente, los fiscales pocas veces reciben objeciones serias a sus hipótesis. Recuérdese que la mayoría de los casos en los sistemas adversariales se solucionan con el beneplácito de los abogados defensores, con negociaciones o admitiendo cargos. Como si ello fuera poco, un fiscal virtuoso no debería acusar a un presunto sospechoso sin estar convencido de la responsabilidad penal del mismo. De esta forma, la ética de la acusación sería inconsistente con la consideración

seria de hipótesis alternativas, o con la misma inocencia del acusado. Así se podría explicar el hecho de que muchas sentencias exculpatorias no se entienden como un triunfo de la justicia, sino como una falla del sistema.

De esta breve identificación del sesgo confirmatorio en los procesos penales se podría inferir que este fenómeno tiene múltiples causas: la psicología humana, el diseño institucional, las expectativas laborales o las concepciones de justicia, entre otras. Yo defiendo que la teoría retórica de la argumentación aumenta los sesgos confirmatorios en los procesos penales. Teniendo esto en cuenta, el siguiente paso de mi argumentación es explicar cómo dicha teoría favorece el sesgo confirmatorio.

3.2. PERSUASIÓN Y SESGO CONFIRMATORIO

Para mostrar que hay una correlación causal entre la teoría retórica y el sesgo confirmatorio usaré los resultados de dos experimentos hechos por O'Brian (2009a) en los que se les pide a los participantes que analicen un expediente de un caso de homicidio, con la expectativa de ser evaluados con respecto a ciertas tareas propias de la actividad que realizan los fiscales. Estas tareas son: "buscar información relevante, interpretarla y construir narrativas coherentes que la integren" (ibíd.: 1025). Aunque los criterios bajo los que los participantes esperaban ser evaluados variaron dependiendo de las condiciones del estudio, dichos experimentos "demostraron que cuando las personas esperan rendir cuentas por sus habilidades de persuadir a otros sobre la culpabilidad de un sospechoso se exacerban los sesgos cognitivos en los procesos de toma de decisión propios de los fiscales" (ibíd.: 1).

Las hipótesis a ser corroboradas por O'Brian fueron tomadas de estudios previos sobre los efectos psicológicos de la rendición de cuentas, que resumiré brevemente. Primero, la rendición de cuentas es definida como "la expectativa, real o hipotética, de que uno será llamado a justificar sus creencias, sentimientos y acciones frente a otros" (Lerner y Tetlock, 1999: 255). Segundo, cuando se dan las condiciones correctas, las rendiciones de cuentas hacen que la persona que espera ser evaluada sea más cuidadosa y objetiva. Sin embargo, hay situaciones en las que la rendición de cuentas afecta la imparcialidad del evaluado (O'Brian, 2009a: 1018). Tercero, a O'Brian le interesó especialmente explorar el impacto que

tienen diferentes tipos de rendición de cuentas en los procesos de búsqueda, interpretación e integración de información fáctica relevante para los procesos de atribución de responsabilidad penal. Así, su hipótesis es que "la forma específica por la que el agente espera ser evaluado afecta el tipo y la intensidad de errores de juicio que comete" (ibíd.: 1020). Esta hipótesis se inspira en dos investigaciones. Por una parte, en un estudio hecho por Itamar SIMONSON y Barry M. STAW (1992) en el que se muestra que las personas que deben rendir cuentas por el resultado de sus procesos de razonamiento tienden a evaluar menos evidencia y a tener creencias más rígidas que las de las personas que son evaluadas por la corrección de sus procesos de razonamiento (ibíd.: 421). Por la otra, un estudio de Philip E. Tetlock, Linda Skitka y Richard Boettger (TETLOCK, SKITKA y BOETTGER, 1989) en el que, primero, se aclara que algunas tareas no son evaluadas por la precisión que se tenga en la descripción de los hechos del pasado, sino por convencer a otros de que se tiene dicha precisión. Luego, se muestra que las personas que adoptan una decisión antes de persuadir a alguien tienden a buscar argumentos que defiendan dicha decisión, en lugar de evaluar diferentes alternativas, como ocurre con las personas que anticipan ser evaluadas por su capacidad de identificar qué decisión es la mejor (ibíd.: 634).

Para todos los efectos relevantes, los dos experimentos de O'BRIAN (2009a) fueron desarrollados de manera similar. Para empezar, los participantes fueron reclutados de un grupo de estudiantes universitarios a los que se les pidió que tomaran decisiones que les permitieran avanzar en una investigación criminal. Ellos fueron divididos aleatoriamente en cuatro grupos, y "a los participantes de tres de estos se les impusieron tres criterios deferentes de rendición de cuentas, mientras que al grupo de control no se le impuso ningún criterio en absoluto" (ibíd.: 1027). Los criterios de rendición de cuentas asignados a cada uno de los grupos fueron el éxito persuadiendo a otros, el uso de un proceso cognitivo correcto y la precisión en el resultado final. Estos estándares de evaluación se hicieron explícitos en las instrucciones que se les dieron a cada grupo. Estos las escucharon a través de una grabación que se reprodujo al comienzo de los experimentos.

Al grupo cuyos participantes deberían rendir cuentas por sus habilidades de persuadir a otros se les dijo:

> Este estudio tiene dos partes, usted está participando en la primera de ellas. Su trabajo es entender qué pasó y construir un breve argumento (que dure aproximadamente 5 minutos) que persuada a un jurado de su posición. Para hacer esto, deberá revisar la evidencia y pensar en el argumento que presentará en una entrevista que será grabada. Esta grabación luego se les presentará a las personas que participarán en la segunda parte de este estudio y será valorado dependiendo de qué tan persuasivo es (ibíd.: 1027).

Las instrucciones dadas al grupo que sería evaluado por usar un proceso cognitivo correcto fueron las siguientes:

> Como usted sabe, el expediente que está a punto de leer hace parte de un caso real. Durante la investigación, el departamento de policía estaba preocupado por la precisión de sus investigaciones, es decir, se preguntó si sus detectives usaban las mejores técnicas de investigación disponibles para encontrar la verdad. El departamento, por lo tanto, contrató un consultor externo para que lo asesorara sobre los procedimientos correctos para maximizar las probabilidades de los detectives para alcanzar la verdad y para revisar el trabajo de los mismos. Este consultor es el mejor en este campo y ha asesorado algunas de las instituciones más grandes en el país, incluido el FBI. Estamos interesados en saber qué tanto sus decisiones investigativas se aproximan a los procesos que él defendió. No estamos tan interesados por los resultados de sus estrategias, sino por la información que tiene en cuenta para formular dichas estrategias. Por lo anterior, al final de este experimento le haremos una breve entrevista sobre sus respuestas y juicios (ibíd.: 1027-1028).

Finalmente, las personas que esperarían ser evaluadas por el mejor resultado posible fueron orientadas de la siguiente forma:

> Estamos interesados en saber si usted puede identificar *quién* cometió el crimen y *cómo* exactamente pasó. Como usted lo sabe, el expediente que está a punto de leer proviene de un caso real. Este caso ya fue juzgado y evaluado por un jurado y un juez. Para medir qué tan acertados son sus juicios sobre lo que pasó los compararemos con las conclusiones del juez y el jurado. Para esto, entrevistaremos a cada uno de ustedes al final del experimento y, posteriormente, analizaremos la grabación de sus entrevistas, para ver si entendieron exitosamente lo que pasó (ibíd.: 1029).

A los miembros del grupo de control simplemente se les pidió que leyeran el expediente y respondieran las preguntas que se les formularían, pero no se mencionó ninguna entrevista o un objetivo particular.

Una vez que cada uno de los grupos recibió sus instrucciones respectivas, se les pidió que leyeran el caso. Tres paquetes de información direccionaron esta lectura. En primer lugar, en las etapas iniciales de la investigación la evidencia sugería que un individuo llamado Bill Briggs era el sospechoso. A modo de ilustración, se afirmó:

> … [que él] tenía antecedentes penales por delitos menores, que había sido despedido por la víctima algunos meses antes del homicidio, que no tenía una coartada para la noche en la que ocurrieron los hechos y que un tendero lo identificó comprando cigarrillos a tres cuadras de la casa de la víctima, cinco minutos antes del momento en el que se cree que el disparo ocurrió (ibíd.: 1024).

Los otros posibles sospechosos, como la ex novia de la víctima, tenían coartadas que los ubicaban en lugares diferentes a la escena del crimen. En segundo lugar, a medida que los participantes avanzaban en su lectura, se les presentó más evidencia confirmando la hipótesis de que Briggs era el sospechoso. Por ejemplo, "la policía encontró un arma sin registro y del mismo calibre de la que mató a la víctima en el apartamento de Briggs, y el tendero identificó a Briggs por segunda vez, en una línea de reconocimiento de personas, como el hombre que compró los cigarrillos la noche del crimen" (ibíd.: 1024). Finalmente, el caso también incluyó evidencia que cuestionaba la hipótesis de que Briggs era el sospechoso: "la policía se dio cuenta de que el tiro de gracia probablemente ocurrió una hora más tarde de lo que inicialmente se creyó", lo que debilitaba las coartadas de otros posibles sospechosos y hacía irrelevante la identificación del tendero. Adicionalmente, en "una segunda inspección a la escena del crimen se encontró en el baño de la víctima cocaína por un valor de mil cien dólares", y la policía también descubrió que "el sobrino de la víctima tenía un serio problema de adicción al juego y que heredaría la mitad de la considerable fortuna de la víctima". Finalmente, "un hombre que no concuerda con la descripción del sospechoso trató de empeñar algunos objetos que coinciden con los que se perdieron de la casa de la víctima la noche del homicidio" (ibíd.: 1025).

Los participantes fueron evaluados en cuatro pruebas. En primer lugar, y antes de conocer el caso hipotético que se les asignaría, se les pidió que expresaran su creencias con respecto a catorce proposiciones relevantes para el caso, en una escala de 1 a 5, en donde 1 era "estoy totalmente de acuerdo" y 5 "estoy totalmente en desacuerdo". Algunas de estas proposiciones eran consistentes con la hipótesis de que Briggs era el culpable y otras eran inconsistentes con ella. Los participantes debieron realizar la misma tarea después de leer el caso. Se esperaba que la variación entre la primera y la segunda vez que realizaran este ejercicio mostraría cómo cambiaron su estado de creencias con respecto a su teoría del caso definitiva. En segundo lugar, asumiendo que "alguien inclinado a confirmar sus hipótesis contra el sospechoso inicial debería recordar más detalles que sean consistentes con la tesis de la culpabilidad y menos con la tesis de la inocencia" (ibíd.: 1026), se les pidió a los participantes que respondieran a un cuestionario de verdadero o falso respecto de cosas que ellos recordaban sobre el caso. En tercer lugar, se pensó que "alguien que busca evidencia de forma sesgada debería también buscar nueva información consistente con la culpabilidad del sospechoso, en lugar de buscar nueva información que soporte otras teorías del caso" (ibíd.: 1026). En consecuencia, los investigadores sugirieron veinte y cinco cosas que la policía debería hacer en el siguiente estadio de la investigación, ocho de las cuales eran consistentes con la hipótesis de que Briggs era el sospechoso. Luego, se les pidió a los participantes que encerraran en un círculo las tres líneas de investigación que, desde su perspectiva, fueran más fructíferas. Finalmente, tratando de identificar las formas en las que los participantes integraron información inconsistente, los investigadores les pidieron que explicaran lo que pensaban sobre la evidencia inconsistente o incompatible con la hipótesis de que Briggs era el culpable.

El resultado de estos experimentos corroboró la hipótesis de que "las personas que esperan ser evaluadas por persuadir a otros de la corrección de su posición son más propensas a los sesgos confirmatorios" (ibíd.: 1028). O'Brian se refiere al primer estudio de la siguiente forma:

> A través de las diferentes pruebas, los participantes de quienes se esperaba que persuadieran a otros sobre su opinión del caso interpretaron la evidencia ambigua o inconsistente en una forma que fuera más consistente con la

> culpabilidad del sospechoso inicial. Comparados con el grupo de control, los participantes que esperaban persuadir a otros interpretaron la evidencia de forma que fuera consistente con su hipótesis inicial. Ellos también propusieron que la policía adelantara más líneas de investigación que se enfocaban en Briggs y cambiaron sus estados de creencias sobre asuntos relevantes en el caso para para reconciliarlos con su teoría de la culpabilidad (ibíd.: 1028-1029).

De forma similar, con respecto al segundo estudio, O'Brian afirma:

> Comparadas con las personas que hacían parte del grupo de control, las personas que esperaban ser juzgadas por sus habilidades de persuadir a otros de sus opiniones mostraron una tendencia a confirmar la culpabilidad del sospechoso inicial seleccionando más estrategias de investigación que se centraran en él, en lugar de otras líneas de investigación alternativas. El mismo patrón emergió de sus recuerdos con respecto a la información consistente con su hipótesis de la culpabilidad, su convergencia de opiniones con respeto a esta hipótesis y sus sugerencias de nuevos patrones de información (ibíd.: 1031).

Si estos resultados son correctos, entonces la teoría retórica de la argumentación conduce al sesgo confirmatorio. Dado que el sesgo confirmatorio es tóxico para el proceso penal, la teoría retórica debe ser reemplazada por una capaz de reducirlo. En la última parte de este escrito presentaré esquemáticamente dicha teoría.

4. La teoría epistemológica de la argumentación y la reducción de los sesgos confirmatorios

Aunque los sesgos confirmatorios son parte de nuestra psicología, deberían generarse estrategias para su reducción. Para ello propongo una teoría epistemológica de la argumentación (Marrero, 2016). Siguiendo una definición estándar, para esta teoría "el típico fin de la argumentación es el conocimiento o la creencia justificada, en sentido epistemológico" (Lumer, 2005: 190). Siendo más precisos, una teoría epistemológica de la argumentación defendería que el *output* típico de la argumentación es el "juicio justificado" y no la "creencia justificada". La razón para esta diferenciación es que "los resultados palpables de las discusiones jurídicas no son creencias privadas sino juicios públicos sobre la culpabilidad o la

inocencia del acusado" (GOLDMAN, 199: 190). Esta teoría sirve mejor el fin de proporcionar justicia sustancial para los individuos porque, desde esta perspectiva, los argumentos deberían estar constituidos por proposiciones verdaderas. Consecuentemente, una teoría epistemológica de la argumentación orientaría mejor al adjudicador con respecto a la verdad de los hechos en contienda. La teoría retórica, por el contrario, no proporciona esta orientación porque la creencia de la que se espera que el adjudicador sea persuadido no está cualificada, es decir, no está sujeta a condiciones posteriores como ser verdadera o altamente probable (LUMER, 2010: 45-46). Como consecuencia, un adjudicador podría ser persuadido de algo que no es verdadero, o de algo que es verdadero por coincidencia. En un caso o en el otro, el adjudicador no conocería lo que se espera que conozca. En la teoría de Schuetz, esta falta de cualificación se hace explícita cuando ella reconoce:

> Algunas de las narrativas construidas por los abogados versan sobre hechos y otras son ficticias. Las historias presentadas por una de las partes en litigio podrían ser falsas, pero estas persuaden a los jueces y jurados porque estas parecen más probables y son más compatibles con las experiencias de los jueces y los jurados que las de las otras partes (SCHUETZ, 2007: 19).

Adicionalmente, una teoría epistemológica de la argumentación jurídica no permitiría los sesgos confirmatorios porque, en términos generales, un agente está justificado en creer algo si no hay proposiciones contrarias que debiliten su creencia. El problema con los sesgos confirmatorios es que, aunque los agentes cognitivos que los padecen tienen información que debilitaría su creencia, ellos la ignoran, la minimizan o creen que no es confiable. El fundamento epistemológico de la argumentación fáctica, por el contrario, considera a los derrotadores cognitivos un elemento central de la argumentación y podría ser útil en la reducción de sesgos cognitivos porque hace al agente responsable de la evidencia contraria a sus proposiciones.

De los resultados de O'Brian podrían surgir algunas objeciones a mi punto de vista. Una teoría epistemológica impondría dos criterios de rendición de cuentas en los que se evaluaría al individuo, ya sea por alcanzar el resultado correcto, es decir, alcanzar el conocimiento, o por aplicar correctamente un proceso cognitivo confiable. Sin embargo, en

los resultados presentados por O'Brian los participantes que esperarían ser evaluados por "alcanzar un resultado correcto o usar el procedimiento correcto no variaron del grupo de control en ninguna de las variables propuestas" (2009a: 1029). Consecuentemente, una teoría epistemológica no mostraría ningún beneficio. Sin embargo, la diferencia de resultados entre el criterio de la persuasión, por una parte, y los del resultado acertado y el uso del procedimiento correcto, por la otra, se puede explicar teniendo en cuenta que al único grupo al que se le crea la presión psicológica propia de la rendición de cuentas es al que se le impone el criterio de la persuasión. Permítaseme exponer mi respuesta paso a paso. Primero, recuérdese que en estos experimentos se entiende la rendición de cuentas como "la experiencia de sentirse presionado a justificar a otros juicios o decisiones" (ibíd.: 1018). Dadas las instrucciones impartidas a los miembros de los grupos en estudio, los únicos que pudieron sentir esta experiencia fueron las personas que esperaban ser juzgadas por su habilidad persuasiva. Mientras que a los miembros de este grupo se les dijo que las personas que participarían en la segunda parte del estudio evaluarían los argumentos construidos dependiendo de qué tan persuasivos eran, a los participantes del grupo del uso del proceso cognitivo apropiado y a los del resultado correcto se les dijo que sus opiniones serían comparadas con las del asesor experto y con las de jueces y jurados, respectivamente. Por lo tanto, estos dos últimos grupos no podían sentir la presión psicológica propia de la rendición de cuentas (ibíd.: 1032).

Una segunda respuesta debería diferenciar entre los criterios del uso del proceso cognitivo correcto y los del resultado apropiado. Con respecto al primero, a los miembros de este grupo se les dice que no es tan importante el resultado de su análisis, sino la manera como lleguen a él. Sin embargo, como lo reconoce O'Brian, no se les explica qué es, ni se les entrena en el uso de un procedimiento cognitivo correcto. Por lo tanto, "ellos podrían trabajar más duro sin hacerlo más inteligentemente" (ibíd.: 1032). Para ilustrar, desde el punto de vista del confiabilismo epistemológico, un proceso cognitivo es una función con *inputs* que tiene creencias como *outputs* (Goldman, 1976: 13). Adicionalmente, la confiabilidad de dichos procesos es la "tendencia de […] producir creencias que son verdaderas en lugar de falsas" (ibíd.: 16). En los experimentos de O'Brian, no se les enseña a los participantes del grupo en cuestión a usar procesos cognitivos confiables,

lo cual puede explicar que no tengan mejores resultados que el grupo de control. De la misma forma, a las personas que esperaban ser evaluadas por la corrección del resultado, es decir, por la verdad en la determinación de los hechos, se les dijo que sus creencias serían comparadas con las de los jueces y jurados que estudiaron el caso con anterioridad. No obstante, el hecho de que un juez o un jurado tome una decisión no implica que su determinación de los hechos en litigio sea verdadera. De hecho, como debería ser claro a esta altura del escrito, los jueces y los jurados también son vulnerables a los sesgos cognitivos. Esto también explicaría por qué este segundo grupo no sería mejor que el de control en la reducción de los sesgos cognitivos.

Las limitaciones propias de los experimentos de O'Brian y las respuestas que ofrezco a las objeciones que se me podrían presentar, explican el título de este ensayo. Mi propósito no ha sido otro que usar algunos experimentos de psicología empírica para construir un argumento en contra de la teoría retórica y no proponer un argumento a favor de la teoría epistemológica. Sin embargo, mi intuición es que los errores que le atribuyo a la teoría retórica podrían ser corregidos por una teoría con fundamento epistemológico. Construir argumentos a favor de esta teoría es una tarea que adelantaré en otros escritos.

REFERENCIAS

ANDERSON, T., SCHUM, D. y TWINING, W., 2005: *Analysis of Evidence*. Cambridge: Cambridge University Press.

ASK, K. y GRANHAG, P. A., 2005: "Motivational sources of confirmation bias in criminal investigations: the need for cognitive closure", *Journal of Investigative Psychology and Offender Profiling*, 2 (1): 43-63.

ASK, K., RABELIUS, A. y GRANHAG, P. A., 2008: "The 'elasticity' of criminal evidence: a moderator of investigator bias", *Applied Cognitive Psychology*, 21 (9): 1245-1259.

ASK, A., REINHARD, M., MARKSTEINER, T. y GRANHAG, P. A., 2011: "Elasticity in evaluations of criminal evidence: exploring the role of cognitive dissonance", *Legal and Criminological Psychology*, 16 (2): 289-306.

BURKE, A., 2006: "Improving prosecutorial decision-making: some lessons of cognitive science", *William & Mary Law Review*, 47: 1588-1633.

BURKE, A., 2007: "Neutralizing cognitive bias: an invitation to prosecutors", *New York University Journal of Law & Liberty*, 2: 512-530.

EERLAND, A. y RASSIN, E., 2012: "Biased evaluation of incriminating and exonerating (non)evidence", *Psychology Crime and Law*, 19 (4): 351-358.

FRANCISCA, F., RAMÓN, A. y MERCEDES, N., 2003: "Cognitive bias and judicial decisions", en VANDEHALLEN, M., VERVAEKR, G., VAN KOPPEN, P. J. y GOETHALS, J. (eds.), *Much ado about Crime*. Bruselas: Uitgeverij Politeia NV, 2878-304.

FINDLEY, K. y SCOTT, M., 2006: "The multiple dimensions of tunnel vision on criminal cases", *Wisconsin Law Review*, 2: 291-397.

GOLDMAN, A. 1976: "What is a justified belief?", en PAPPAS, G. (ed.), *Justification and Knowledge*. Dordrecht: D. Reidel, 1-23.

GOLDMAN, A., 2003: "Simple heuristics and legal evidence", *Law, Probability and Risk*. 2: 215-226.

LERNER, S. J. y TETLOCK, P. E., 1999: "Accounting for the effects of accountability", *Psychological Bulletin*, 25 (2): 255-275.

LERNER, S. J. y TETLOCK, P. E., 2003: "Bridging individual, interpersonal, and institutional approaches of judgement and decision making: the impact of accountability in cognitive bias", en SCHNEIDER, S. L. y SHANTEAU, J. (eds.), *Emerging Perspectives on Judgement and Decision Research*. Cambridge: Cambridge University Press, 431-457.

LUMER, C., 2005: "The epistemological theory of argument. How and why?", *Informal Logic*, 25 (3): 189-212.

LUMER, C., 2010: "Pragma-dialectics and the function of argumentation", *Argumentation*, 24 (1): 41-69.

Marksterner, T., Ask, K., Reinhard, M. y Granhag, P. A., 2011: "Asymmetrical skepticism towards criminal evidence: the role of goal- and belief-consistency", *Applied Cognitive Psychology*, 25 (4): 541-547.

Marrero, D., 2016: "An epistemological theory of argumentation for adversarial legal proceedings", *Informal Logic*, 36 (3): 288-308.

O'Brien, B., 2009a: "A recipe for bias: an empirical look at the interplay between institutional incentives and bounded rationality in prosecutorial decision making", *Missouri Law Review*, 74 (4): 999-1050.

O'Brien, B., 2009b: "Prime suspect: an examination of factors that aggravate and counteract confirmation bias in criminal investigations", *Psychology, Policy and Law*, 15 (4): 315-334.

Perelman, C., 1980: "Law and rhetoric", en *Justice, Law and Argument. Essays on Moral and Legal Reasoning.* Dordrecht: D. Reidel Publishing Co., 120-135.

Rymond, N., 1998: "Confirmation bias: a ubiquitous phenomenon in many guises", *Review of General Psychology*, 2 (2): 175-220.

Rassin, E., Eerland, A. y Kuijpers, I., 2010: "Let's find the evidence: An analogue study of confirmation bias in criminal investigations", *Journal of Investigative Psychology & Offender Profiling*, 7 (3): 231-246.

Schuetz, J., 1994: *The Logic of Women on Trial. Case Studies of Popular American Trials.* Carbondale: Southern Illinois Evicersity Press.

Schuetz, J., 2007: *Communicating the Law. Lessons from Landmark Cases.* Long Grove: Waveland Press.

Simonson, I. y Staw, B. M., 1992: "Deescalation strategies: a comparison of techniques for reducing commitment to losing courses of action", *Journal of Applied Psychology*, 77 (4): 419-426.

Tetlock, P. E., Skitka, L. y Boettger, R., 1989: "Social and cognitive strategies for coping with accountability: conformity, complexity, and bolstering", *Journal of Personality and Social Psychology*, 57 (4): 632-640.

VIII.
Entre el 'hombre promedio' y el 'hombre ideal': el criterio de razonabilidad y su aplicación al agente moral en el razonamiento cotidiano, jurídico y psicológico

MARÍA NATALIA ZAVADIVKER*

INTRODUCCIÓN: EL PROBLEMA DE LA CIRCULARIDAD ENTRE DESCRIPCIÓN Y NORMATIVIDAD APLICADO A LA REPRESENTACIÓN DEL AGENTE MORAL

En este capítulo me propongo estudiar uno de los aspectos claves implicados en el abordaje de cuestiones ligadas al ámbito legal y jurídico desde la perspectiva de la psicología experimental. La mayor dificultad que revisten los intentos de que una ciencia empírica (la psicología moral) proporcione aportes relevantes a disciplinas y ámbitos de aplicación normativas (la legislación, la jurisprudencia, la política, etc.) reside precisamente en la posibilidad de delimitar la frontera, y a su vez establecer interacciones dialécticas, entre el plano descriptivo y el normativo. En este caso abordaré un aspecto sumamente relevante ligado a dicho conflicto: las dificultades a la hora de forjar una representación del agente moral, que implican una confusión entre el modo como este es percibido *de hecho* y el modo como creemos que *debería ser*. En este caso no se trata, en principio, del problema de si es posible derivar normas a partir de cuestiones de hecho (si bien también es relevante la cuestión de si es posible o deseable que las decisiones en materia legal y jurídica se nutran del conocimiento científico de la naturaleza humana). Lo que pretendo analizar es el asunto en un sentido inverso: la cuestión de la dimensión normativa implícita en la tarea, aparentemente neutral, de describir el comportamiento y la cognición moral humanos.

La comprensión implícita de cómo opera y cómo creemos que debería operar el agente moral resulta, como procuraremos demostrarlo más adelante, absolutamente relevante en muchos dominios. En el dominio de la investigación psicológica, cuyo objetivo es abordar al agente moral en su

---

* Doctora en Filosofía de la Universidad Nacional de Tucumán (Argentina). Profesora asociada de Epistemología de la Facultad de Bioquímica, Química y Farmacia (FBQF) y profesora adjunta de la misma materia en la Facultad de Educación Física de la Universidad Nacional de Tucumán. Investigadora asistente del Consejo Nacional de Investigaciones Científicas y Técnicas – CONICET (Argentina). Coordinadora de la Comisión de Bioética de la FBQF e integrante del Comité de Ética de la Investigación (UNT/CONICET).

dimensión empírica, se vuelve necesario sin embargo que los investigadores fijen ciertos parámetros de 'normalidad' asociados a estándares implícitos de 'corrección moral' a los fines de evaluar el comportamiento natural de los sujetos tomando dichos parámetros como punto de referencia. Por ejemplo, la detección de sesgos y falacias en el razonamiento moral humano no sería posible si no presupusiéramos de antemano la existencia de un razonamiento moral perfectamente racional, de modo tal que los sesgos serían meras desviaciones o distorsiones que violarían ciertas normas implícitamente asumidas. Lo mismo cabe decir en relación con la evaluación de comportamientos, hábitos, creencias, juicios de valor, etc.

Pero dichos parámetros implícitos en el criterio de los investigadores, si son a su vez subsidiarios de mecanismos psicológicos propios de la especie humana en su conjunto, también deberían operar cuando cualquier persona juzga el comportamiento de otra (por ejemplo, cuando un juez o un jurado evalúan el comportamiento de un agente para juzgar su inocencia o culpabilidad –o factores presuntamente atenuantes o agravantes–, o cuando cualquier individuo interpreta y juzga moralmente el comportamiento, motivaciones, creencias y valoraciones de otros). Los psicólogos morales pretenden desentrañar esos mecanismos psicológicos subyacentes a la cognición y el comportamiento moral mediante un abordaje empírico, utilizando herramientas experimentales (tests conductuales y basados en imágenes cerebrales, etc.), pero para tal fin deben necesariamente aplicar sus propios criterios implícitos de "corrección moral" o ajuste del comportamiento a ciertos parámetros, como un referente con el cual medir y comparar el desempeño observado en los sujetos experimentales. En otras palabras, presuponen una normatividad implícita, del mismo modo que la presupone también cualquier sujeto cuando evalúa el comportamiento, valoraciones y razonamientos morales de otros. Dicha normatividad es, a su vez, parte de lo que los psicólogos morales quieren descubrir empíricamente, lo que nos remite al problema de la inevitable circularidad entre descripción y normatividad.

1. Teorías estadísticas, prescriptivas e híbridas

Dicho esto, la pregunta que nos formularemos a continuación es la siguiente: suponiendo que estos estándares implícitos sean concebidos como

criterios de "normalidad", vale decir, que existan presuposiciones a priori acerca de cuáles son (o deberían ser) los comportamientos, juicios, razonamientos etc. 'normales' en los seres humanos, ¿cómo debería entenderse dicho criterio? Existen dos grupos de teorías alternativas que procuran responder a esta pregunta:

– Para las teorías estadísticas, la 'normalidad' es concebida en un sentido puramente descriptivo como la pauta conductual (incluyendo por tal comportamientos, creencias, valoraciones, etc.) sustentada por la mayoría de las personas, o, dicho de otro modo, como la respuesta conductual que aparece con mayor frecuencia estadística. Si tal fuera el caso, las investigaciones experimentales en psicología moral contribuyen a proporcionar precisamente este dato, ya que aportan evidencias sobre el porcentaje de individuos que exhibe tal o cual comportamiento, toma tal o cual decisión o valora tal o cual situación moralmente relevante frente a ciertos estímulos.

– Las teorías prescriptivistas, en cambio, argumentan que en los hechos no solemos (ni deberíamos hacerlo) tomar la mera frecuencia estadística (el comportamiento y las valoraciones más comunes y extendidas) como un parámetro de lo que es o debería ser 'normal'. Por ejemplo, la evidencia empírica de que el 87% de los conductores de automóvil revisa su celular mientras está conduciendo no debería ser tomada como un criterio de 'normalidad' y por ende considerada como un atenuante a la hora de mitigar la condena de un automovilista que provocó un accidente por conducir mirando su celular; o la evidencia de que durante la Alemania nazi el 98% de los alemanes tenía odio racial contra los judíos (de modo tal que el racismo era lo 'normal' en ese contexto histórico y cultural) no debería ser un argumento para reducir la condena a los responsables del genocidio nazi. En tal sentido, para las teorías prescriptivistas, la normalidad debería ser concebida como el ajuste del comportamiento, razonamiento o valoraciones morales a alguna 'norma', cuyo incumplimiento debería ser tomado como una suerte de desviación (y en el límite como una patología, no solo aplicable a individuos particulares, ya que incluso podría tratarse de patologías colectivas de origen social, como en los ejemplos dados). En ese sentido, para este grupo de teorías el criterio de normalidad posee una clara dimensión normativa.

Dicho criterio prescriptivista a su vez puede ser concebido en dos sentidos. El segundo de ellos, plasmado en el conjunto de teorías que

Kevin Tobia denomina *híbridas* (Tobia, 2018: 305), pretende precisamente rescatar el punto medio entre descripción y normatividad, y aparece como el candidato más promisorio tanto para establecer algún parámetro normativo que nos permita juzgar adecuadamente los comportamientos morales relevantes en la esfera del derecho y la jurisprudencia como para representar los juicios morales que de hecho realiza la mayoría de las personas cuando evalúa comportamientos como 'normales'.

– El primer sentido apela a un criterio 'ideal' de corrección racional, que aun cuando no refleje el razonamiento real de la mayoría de las personas, es percibido implícitamente como la forma de razonamiento o comportamiento óptimo de un agente moral puramente racional en condiciones ideales. Esta perspectiva (coincidente con la lógica subyacente a los modelos de optimización utilizados en biología evolutiva, economía, etc.) implica fijar una serie de estándares normativos ideales para luego comprobar en qué medida el desempeño promedio de seres humanos reales se acerca o se aleja del desempeño 'óptimo'. Dichos criterios de optimización, que implican distintos tipos de 'racionalidad', también pueden a su vez variar dependiendo de la concepción moral implícita que se sustente. Por ejemplo, en una perspectiva deontologista de raíz kantiana lo correcto sería, desde el punto de vista conductual, el cumplimiento del deber moral, cumplimiento motivado a su vez por principios que se da a sí misma la razón y no por inclinaciones como sentimientos, intereses, deseos o preferencias. Desde esta misma perspectiva, las evaluaciones y razonamientos morales correctos serían, por ejemplo, aquellos que juzgan las intenciones del agente más que los resultados de sus acciones, o que procuran realizar evaluaciones objetivas e imparciales evitando sesgos de todo tipo, o que se guían por principios de justicia intentando alcanzar un equilibrio entre intereses contrapuestos, etc. En cambio, desde una perspectiva utilitarista o consecuencialista, que pone el énfasis en un tipo de racionalidad más bien estratégica e instrumental, el agente moral ideal será aquel que obre, juzgue otros comportamientos o guíe su razonamiento hacia las decisiones que maximicen el bienestar de la mayoría de los sujetos implicados, vale decir, que apunten al mejor resultado posible como fruto de cálculos utilitarios (balances entre costes y beneficios). Se trata de un esquema similar al agente racional-egoísta propio de los modelos clásicos en economía conductual, pero cuyo objetivo no es maximizar el propio beneficio (autointerés) sino

alcanzar el mayor bien, entendiendo por tal los mayores beneficios y los menores perjuicios posibles para la mayoría de los sujetos implicados. A diferencia del criterio deontologista, valora los resultados de una acción como más relevantes que las intenciones del agente. Retomando la cuestión de la circularidad entre las concepciones teóricas y lo que se "descubre" mediante el uso de herramientas empíricas, en la actualidad se han encontrado correlatos neuronales (en regiones diferenciadas del cerebro) de mecanismos psicológicos asociados tanto al criterio deontologista como al utilitarista (Greene, Nystrom, Engell, Darley y Cohen, 2004), lo que aportaría evidencias de que el razonamiento moral humano se sustentaría en la conjunción entre ambas perspectivas (vale decir, las evaluaciones morales suelen tomar en cuenta tanto las intenciones del agente como los resultados de la acción).

Kevin Tobia (2018: 302) menciona tres grupos de teorías prescriptivistas que implican distintos tipos de racionalidades: el primero coincide con las ya mencionadas teorías de la optimización que identifican la racionalidad con estándares de eficacia coste-beneficio en aras de la maximización del bienestar. El segundo grupo en realidad implica la asunción de una relativización del criterio de racionalidad, al concebirlo como el conjunto de valores y supuestos normativos vigentes en una determinada comunidad y durante un determinado período histórico (una suerte de contrato implícito entre los miembros de una comunidad en virtud del cual se fijan los estándares normativos básicos que regirán para dicho grupo). Volveremos sobre este criterio más adelante, cuando hagamos referencia a las teorías híbridas e introduzcamos el concepto de 'razonabilidad'. Baste mencionar por ahora que dicho criterio podría entenderse a su vez como una mera descripción metaética del conjunto de normas y valores que rigen implícitamente en una determinada comunidad, pero no parece tener en sí mismo una pretensión normativa. En el tercer grupo, Tobia incluye aquellas teorías en las que el criterio de racionalidad no es la eficacia coste-beneficio sino alguna otra noción prescriptiva a la que se le adjudica un valor moral objetivo (racionalidad como perfeccionamiento de las virtudes morales, como promoción de la libertad, como justificación, etc.).

– El segundo modo de concebir un criterio de 'normalidad' es el propuesto y analizado por Kevin Tobia en su trabajo "How people judge what is reasonable", publicado en *Alabama Law Review* en 2018, en el cual apela

a un conjunto de teorías *híbridas*, vale decir, teorías que buscan alcanzar un punto medio entre las evidencias estadísticas y los ideales normativos. Al referirnos a las teorías prescriptivistas hemos utilizado el concepto de 'racionalidad'[1]. Las teorías híbridas, en cambio, apelan a la noción de 'razonabilidad' como presunto parámetro implícito que la gente suele considerar cuando juzga el desempeño moral de un agente como 'normal'. Tobia parece asimilar el criterio de normalidad al de *razonabilidad*[2], y propone la hipótesis, que luego somete a contrastación experimental, de que el criterio implícito de razonabilidad que manejan espontáneamente las personas es un 'híbrido' (o una instancia intermedia) entre un criterio puramente estadístico (lo que sucede en los hechos, o lo que la mayoría de la gente *hace*) y un criterio prescriptivo (el ideal normativo, o lo que *sería ideal que la gente hiciese*). En otras palabras, lo 'razonable' sería que

---

1 En realidad, Tobia utiliza también el concepto de 'razonabilidad', y no el de 'racionalidad', cuando hace referencia a las teorías prescriptivistas. Sin embargo, si él mismo parece definir la 'razonabilidad' esperable en un comportamiento o evaluación moral como el resultado de un balance entre lo que la gente espera que suceda *de hecho* (comportamiento efectivo esperable de sujetos, grupos, corporaciones, etc.) y lo que idealmente quisiera que suceda, o piensa que 'debería suceder' en condiciones ideales, entonces me parece que debería reservar el término 'razonabilidad' solo para las teorías híbridas, y usar el término 'racionalidad' (que es el que de hecho suele usarse) para las teorías prescriptivistas. De todos modos la frontera entre racionalidad (desempeño ideal u óptimo) y razonabilidad (desempeño dentro de límites aceptables) no justifica una división tajante entre las teorías, ya que algunos modelos de optimización pueden establecer también estándares o márgenes de razonabilidad en el comportamiento (por ejemplo, fijar cierto rango de valores en los que un comportamiento sería aceptable porque su costo no está por encima del posible beneficio a obtener).

2 En realidad el autor expone tres variantes híbridas de teorías de la razonabilidad: la del "ideal corregido" (*corrected ideal*), según la cual la razonabilidad se determina ajustando los estándares ideales a la realidad estadística; la del "promedio corregido" (*corrected average*), que, por el contrario, concibe a la razonabilidad como el ajuste de los resultados estadísticos a ciertos estándares prescriptivos; y finalmente, el "estándar combinado" (*blended standard*), según el cual la razonabilidad puede entenderse mejor como un estándar que trata los hechos estadísticos y prescriptivos como los dos indicadores relevantes de razonabilidad. Según el autor (Tobia, 2018: 330), para este último punto de vista la razonabilidad podría ser interpretada como equivalente a lo que suele entenderse por "normalidad".

los agentes morales se comporten dentro de un rango o margen aceptable teniendo en cuenta sus posibilidades fácticas *reales* (las características propias de 'naturaleza humana', que presuntamente se manifestarían en el comportamiento del individuo 'promedio'), y ciertos estándares normativos que deben satisfacerse para que la vida en sociedad sea posible. Por ende, también sería razonable que quienes deben evaluar el desempeño moral de otros agentes (en el ámbito jurídico, por ejemplo) tengan en cuenta estos mismos márgenes de aceptabilidad o razonabilidad, en lugar de pretender un comportamiento ideal no compatible con las evidencias fácticas (desempeño real del individuo promedio). Y en un nivel metaético, también los psicólogos morales apelarían de hecho (y deberían apelar) a este criterio de razonabilidad como un parámetro estándar para juzgar la 'normalidad' en el comportamiento estadístico de los sujetos experimentales.

1.1. El 'hombre común': ¿mayoría estadística o representación ideal de nuestros criterios de razonabilidad?

¿Quiénes fijan los 'estándares de razonabilidad' que a su vez serían representativos de la 'normalidad'? ¿La mayoría estadística o un tipo ideal de hombre o ciudadano que, si bien es 'común y corriente', sustenta algún criterio de razonabilidad que los demás podríamos juzgar, como resultado de un acto reflexivo, como 'razonablemente aceptable'?, ¿dependen dichos criterios de razonabilidad, a su vez, de parámetros socioculturalmente dependientes?, ¿son representativos de la 'naturaleza humana' en general, o subsidiarios de cierta civilización (por ejemplo, la occidental y cristiana), de cierta clase social (por ejemplo, la clase media instruida), de cierto sesgo de género o de cualquier otro sesgo que agrupe o represente a cierto *tipo* de ser humano? Al hablar de 'mayoría estadística', ¿debemos referirnos a la mayoría de los sujetos dentro de un entorno sociocultural o un período histórico determinado?, ¿o a la mayoría de la población mundial en general a lo largo de toda la historia de la humanidad? Como procuraremos demostrarlo, el problema de los estándares ideales u óptimos de racionalidad que presentarían las teorías netamente prescriptivistas es trasladable también a las teorías híbridas de la razonabilidad.

Tobia realiza un extenso y detallado análisis teórico en el que procura mostrar que la historia del concepto de 'persona razonable' (para referirse a la cual apela a un sucedáneo metafórico: 'el hombre del ómnibus de Clapham' (*Man of the Clapham omnibus*) sugiere esta misma concepción híbrida de razonabilidad (a medio camino entre el dato estadístico y la prescriptividad) que el autor defiende. El 'hombre del ómnibus de Clapham' es una noción metafórica que alude a un hipotético ciudadano ordinario razonable, utilizada por los tribunales ingleses en aquellos casos en los que se debe decidir si una de las partes involucradas en un conflicto (por ejemplo, una acción civil por negligencia) actuó como lo haría una 'persona razonable'. De allí que varios académicos y juristas a menudo caractericen a esta ficción legal como el antecedente histórico del concepto de razonabilidad. Sin embargo, semejante noción da lugar a controversias, ya que, tal como lo señala indirectamente Walter Bagehot en el tratado de constitución inglesa de 1873, el hombre común no es representativo de las clases más cultas y educadas –las cuales en toda sociedad conforman una élite intelectual, vale decir, no son representativas del pensamiento de la mayoría–. Por el contrario, el 'hombre del ómnibus de Clapham' representaría a la masa ordinaria de la población, con cierto nivel de educación pero por lo general poco reflexiva y, por ende, sensible al 'poder despótico' de la opinión pública. El criterio de Bagehot, como señala Tobia, no parece coincidir con la idea de que el hombre común puede ser concebido como la mera encarnación de un ideal normativo estándar. Más bien, el hombre de Clapham sería representativo, de acuerdo a dicho criterio, del conjunto de normas y valores vigentes en una determinada comunidad durante determinado período histórico, o bien de la opinión de la mayoría estadística, o de algún punto intermedio entre ambas nociones (Tobia, 2018: 334).

La sentencia de Bagehot[3] también sugiere a las claras un problema recurrente en los estudios sobre razonabilidad: el de la identificación del

---

3 *"The middle classes—the ordinary majority of educated men—are in the present day the despotic power in England. "Public opinion," now-a-days,"is the opinion of the bald-headed man at the back of the omnibus." It is not the opinion of the aristocratical classes as such; or of the most educated or refined classes as such; it is simply the opinion of the ordinary mass of educated, but still commonplace mankind*". Bagehot (1873: 325-326).

hombre ordinario con una clase social determinada (la clase mayoritaria, y al mismo tiempo sensible al pensamiento de masas). Por otra parte, si bien en el caso de Inglaterra, país donde se originaron estos conceptos, la clase media probablemente coincide con la mayoría estadística de los ciudadanos, en los países en desarrollo con mayor desigualdad social la mayoría estadística suele coincidir con los sectores más desposeídos. Sin embargo, es muy claro que esa mayoría numérica no es la que define o determina en los hechos los valores y criterios ideales de razonabilidad vigentes –al menos aquellos relevantes en el ámbito de las decisiones jurídicas–, que siguen siendo los de la clase media instruida. En otras palabras, se asume que el hombre de Clapham, aunque representa al hombre ordinario (en Argentina solemos hablar del 'ciudadano de a pie'), hace referencia a un hombre con el nivel de educación y civilidad necesarios para que su opinión, pensamientos y comportamientos puedan ser juzgados como 'razonables'. Por otra parte, esto aparentemente estaría indicando que la noción de 'hombre común' no coincide con nociones tales como 'lo que el hombre es por naturaleza', o con el supuesto de que tales criterios de razonabilidad dependerían meramente de predisposiciones biológicas de la especie (y los estudios experimentales en psicología moral apuntan más bien a la detección de comportamientos, juicios y razonamientos morales que, de encontrarse de manera recurrente en un amplio grupo de personas, se asume que dependerían de patrones psicológicos universales propios de la arquitectura cognitiva humana). El hombre común sería, por el contrario, el producto de un proceso de socialización que recibió cierto tipo de educación y asimiló ciertos valores, dependientes no solo de su entorno cultural sino también de su clase social, lo que abre la puerta al ingreso de teorías de corte más sociologista según las cuales el criterio del hombre común es representativo de ciertos sectores (e incluso estaría ligado a ciertos intereses de clase, a cierto tipo de civilización que impuso sus propios valores a las demás, etc.), que el hombre común reproduce los discursos y mandatos de los grupos hegemónicos, etc. De allí que una de las líneas de investigación relevantes dentro de la psicología moral experimental sean los estudios multiculturales y con sujetos de diferentes clases sociales y niveles socioeducativos que apunten a detectar, por ejemplo, si existen criterios y estándares básicos de razonabilidad compartidos por cualquier sujeto más allá de su impronta cultural, su estrato social, su nivel

de escolarización, etc. Sin embargo, otra discusión relevante es la de si los datos empíricos sobre el comportamiento y las opiniones de la mayoría de las personas pueden afectar nuestros criterios normativos de razonabilidad (o, en términos generales, aquellos criterios que surgen de la reflexión teorética y el análisis conceptual en profundidad de filósofos, psicólogos, juristas, etc.), o bien si, como señalan los autores prescriptivistas, lo que opina la mayoría estadística de las personas en el plano fáctico debería ser irrelevante y no afectar nuestros estándares normativos.

Pero el problema sigue siendo: ¿quién tiene derecho a fijar dichos estándares? Para responder a esta pregunta resulta útil apelar a las teorías duales de la cognición. La psicología moral estándar en la actualidad pone más el acento en la detección de juicios morales espontáneos y emocionalmente mediados que se disparan de manera más o menos automática en respuesta a ciertos estímulos. Sin embargo, los dilemas que deben resolverse en el terreno jurídico son precisamente situaciones complejas que requieren de la ponderación de múltiples variables, lo que precisa a su vez de la apelación a procesos de razonamiento lentos y reflexivos, que no suelen ser los que realizan espontáneamente las personas, justamente porque no tienen un entrenamiento especial en el tema ni en su vida diaria se dedican específicamente a resolver problemas jurídicos o tomar decisiones en materia legal y política. El pensamiento rápido y espontáneo, aquel que no requiere de un esfuerzo lento y reflexivo, suele apelar a prejuicios y generalizaciones burdas, a la apropiación del discurso social vigente en los medios de comunicación y la opinión pública, a la repetición de consignas y clichés de determinados grupos de referencia con los que se identifica el usuario, etc. En tal sentido, las intuiciones espontáneas de la gente claramente no parecieran ser el parámetro correcto para juzgar y tomar decisiones sobre asuntos complejos en los que esté en juego el futuro de terceros. Ahora bien, Knobe y Nichols, considerados los padres de la filosofía experimental (una nueva línea de investigación que procura capturar intuiciones filosóficas *folk* en el pensamiento espontáneo de las personas –y los procesos psicológicos subyacentes–, y cuyo objeto de estudio se solapa en gran medida con el de la psicología moral experimental), señalan en su manifiesto de 2008 (KNOBE y NICHOLS, 2008) que el enfoque de la filosofía experimental (*X-PHI*) toma como guía orientativa la hipótesis heurística de que en la base de elaboraciones conceptuales más complejas

(fruto de procesos de razonamiento lentos y reflexivos, como la reflexión filosófica o jurídica) podría haber ciertas intuiciones espontáneas de carácter universal –por ejemplo, sobre asuntos tales como determinismo y libre albedrío, justicia retributiva y distributiva, intenciones y resultados de una acción, etc.– que operarían, por así decirlo, como condición de posibilidad de las elaboraciones abstractas posteriores. Y dicha hipótesis es precisamente lo que las investigaciones en filosofía y psicología moral experimental procuran contrastar experimentalmente mediante estudios multiculturales y sociodemográficos.

Este tema nos remite a su vez al enfoque relativista del criterio de racionalidad sustentado por una de las teorías prescriptivistas mencionadas por Tobia (y expuesta anteriormente) según la cual los estándares de racionalidad de un comportamiento pueden ser socioculturalmente dependientes y estar basados en acuerdos tácitos entre los miembros de una comunidad sobre la base de los valores vigentes en la misma. Dicha perspectiva (aplicada a la racionalidad) es también aplicable a los juicios sobre estándares de razonabilidad, ya que en diferentes culturas y momentos históricos los miembros de diversas comunidades pueden sustentar ciertos parámetros ligados al rango de razonabilidad de ciertos actos (por ejemplo, el criterio sobre la vestimenta aceptable de una mujer en la playa –margen razonable de cuerpo al descubierto que sería aceptable que exhiba en público a fin de resguardar su "decoro"– ha variado enormemente en los últimos 50 o 60 años). Si bien en este caso el problema del relativismo y la sujeción histórica y sociocultural de las normas vigentes es abordado desde una perspectiva prescriptivista (el deber-ser se definiría en función de lo que un grupo determinado juzga correcto en un determinado contexto), dichas normas implícitas o explícitas son a su vez un dato empírico que puede ser recabado mediante procedimientos estadísticos (por ejemplo, la detección del porcentaje de individuos de una misma comunidad, o incluso de un mismo grupo social, sub-cultura, profesión, etc., que comparte los mismos estándares de razonabilidad en relación con un asunto determinado). Vale decir, como ya lo mencionamos anteriormente, que si bien los parámetros de razonabilidad fijados por una comunidad son normativos, un análisis metaético tendiente a detectar la existencia de hecho de los mismos en diversas comunidades sería meramente descriptivo, pero a su vez dicho análisis se guiaría implícitamente por algún estándar normativo desde el

cual juzgaría las creencias morales en diferentes épocas y culturas, lo que nos remite nuevamente al problema de la circularidad entre descripción y normatividad.

2. La importancia de fijar parámetros de razonabilidad en el ámbito legal y jurídico: el experimento de Tobia

Por otra parte, la noción de razonabilidad como criterio de normalidad resulta sumamente relevante en los terrenos legal y jurídico, ya que dichos ámbitos sustentan básicamente una moral minimalista, vale decir, no requieren de los ciudadanos un comportamiento moralmente perfecto o que aspire a la perfección, solo exigen que al menos los déficits en el desempeño moral de los sujetos se manejen dentro de límites razonables, o, dicho de otro modo, cumplan con estándares mínimos de razonabilidad. Si bien existen transgresiones claramente especificadas en los códigos civil y penal en las que simplemente se aplica la ley ante una falta y los criterios de razonabilidad no tienen mayor relevancia, hay situaciones más ambiguas en donde sí es relevante la discusión sobre los márgenes de razonabilidad de las acciones. Tobia apela, entre otras referencias, a los casos de daños accidentales, vale decir, cuando un agente provoca alguna catástrofe o efecto negativo no de forma intencional sino por accidente. Da el ejemplo de un pintor que mientras está pintando una casa deja la lata de pintura abierta y el perro de la familia se come el contenido y muere intoxicado. En tal caso, si un jurado debiera decidir hasta dónde el agente es culpable, una vía adecuada (que de hecho es la que suele utilizarse) sería juzgar si los recaudos que tomó estaban dentro de límites razonables, o si su comportamiento fue realmente negligente por no haber tomado las precauciones necesarias dadas las circunstancias.

Tobia realizó tres experimentos cuyo objetivo general era comprobar si los participantes tendían a evaluar los márgenes de razonabilidad de un gran número de acciones y situaciones mediante la asignación de valores cuantitativos que fueran aproximadamente intermedios entre el valor asignado a situaciones *reales* y a situaciones *ideales.* Con bastante ingenio apeló a ejemplos (que se detallarán a continuación) a los que se les pudiese asignar un valor numérico, a los fines de obtener indicadores cuantitativos

objetivamente mensurables. Los participantes fueron reclutados a través de la página web Amazon's Mechanical Turk (un sitio de internet perteneciente a Amazon que brinda a sujetos particulares la opción de realizar tareas que requieren inteligencia humana –vale decir, no realizables por una máquina– a cambio de una compensación económica). En el primer experimento (en el que participaron 84 sujetos), solicitó que estos indicaran los valores que consideraban reales, razonables e ideales en relación con una serie de ítems ligados a la vida cotidiana (por ejemplo, la cantidad de calorías que ingiere una persona por día, el número razonable de libros que una persona lee en un año, el número de horas de TV que mira una persona diariamente, el número de mentiras que una persona dice en una semana, etc.). En el segundo experimento (en el que participaron 59 sujetos), la tarea solicitada y los ítems presentados eran similares (por ejemplo, cantidad razonable de dinero que una persona oculta en su declaración impositiva), pero el test contenía una introducción en la que se procuraba enmarcar las preguntas en un contexto legal, ya que estaba precedido por un texto en el que se les advertía a los participantes que debían juzgar cantidades legalmente razonables (además del comportamiento promedio y el ideal) en una serie de situaciones judicializables, y se les pedía que imaginaran que emitían su veredicto en el marco de un proceso judicial (como si fueran parte del jurado). El texto es bastante explicativo y hasta se brindan ejemplos a los participantes (como el anteriormente citado de si un pintor tomó o no los recaudos razonables para evitar el accidente del perro que se comió la pintura). En el tercer experimento los ítems incluidos están más directamente relacionados con el contexto legal (por ejemplo: días de plazo para aceptar la firma de un contrato de negocios, semanas para devolver un producto ordenado *online*, costo adicional en un contrato para la construcción de un edificio, tiempo que un acusado de un delito debe esperar en prisión preventiva antes de ser juzgado, tiempo que debería demorarse un fallo judicial, tasa de interés de un préstamo, etc.), pero uno de los grupos recibió solo las instrucciones dadas en el experimento 1, mientras que el otro recibió también las instrucciones del experimento 2 (que hacen referencia al contexto legal y las condiciones jurídicas en las que debe enmarcarse la evaluación). En síntesis, en los experimentos 1 y 2 había algunos ítems más interpretables desde un marco legal y otros menos (o directamente no interpretables desde dicho marco), mientras el

tercer experimento testeaba los juicios de razonabilidad en condiciones más estrictamente legales. El autor seleccionó ejemplos muy variados que pudieran ser representativos de un amplio espectro de situaciones correspondientes a diferentes dominios legales –desde el plazo para aceptar un contrato de negocios hasta el número razonable de eventos ruidosos que los vecinos de un club de fútbol deberían soportar–.

Un poco más específicamente, el estudio apuntaba a examinar si tanto los juicios de razonabilidad comunes como los pertenecientes al dominio legal toman más en consideración el comportamiento promedio o el comportamiento ideal, vale decir, si el valor "razonable" indicado está más cerca de lo que los participantes indicaron como valor promedio, como valor ideal, o si es un valor intermedio entre ambos. No nos detendremos en los detalles técnicos del estudio ya que son bastante complejos; baste decir que, en términos generales, se encontró que los valores de razonabilidad elegidos *coincidían de un modo muy significativo con el promedio entre los valores que los sujetos consideraban reales y los que consideraban ideales*, esto es, en muchos casos estaban casi justo en el medio.

Resulta interesante el intento de Tobia de disociar experimentalmente el criterio de razonabilidad asociado a comportamientos de la vida cotidiana respecto de aquel que tiene relevancia en el ámbito jurídico. Para tal fin, como decíamos, recurrió a dos estrategias: 1.º) presentar ejemplos similares en dos condiciones experimentales pero introduciendo en una de ellas una contextualización jurídica en el marco de la cual debía tomarse la decisión, y 2.º) presentar, en experimentos separados, dos conjuntos de ítems diferenciados (uno sobre el margen de razonabilidad en situaciones y comportamientos cotidianos, y el otro sobre el mismo margen aplicado a situaciones judicializables). Mientras en los ítems que hacían referencia a situaciones de la vida cotidiana utilizó solo el término 'razonable', para los ítems con implicancias legales y jurídicas (o cuando solicitó que los ítems ligados a la vida cotidiana se interpretaran desde una perspectiva jurídica) utilizó el término 'legalmente razonable'. Los resultados fueron los mismos en ambos casos: el valor razonable se situó casi en todos los casos en el punto medio prácticamente exacto entre los valores promedios e ideales, lo que estaría indicando que solemos tomar el rango de razonabilidad aplicado a cualquier evaluación de comportamientos y situaciones en

la vida cotidiana como un parámetro adecuado para juzgar los márgenes de acción razonables en el ámbito jurídico.

2.1. HACIA UNA EXPLICACIÓN EVOLUCIONISTA DEL HALLAZGO EXPERIMENTAL DE TOBIA

Tobia se limitó a demostrar experimentalmente que el criterio de razonabilidad sustentado por la mayoría de las personas se encuentra a medio camino entre lo que suele suceder de hecho y una situación presuntamente ideal (la cual sería algo así como el reflejo de nuestras más altas expectativas o deseos), pero no fue más allá. En lo que sigue propondré una posible explicación –desde la óptica de la psicología evolucionista, a la que se suman los aportes de la teoría de juegos– de los mecanismos psicológicos subyacentes a la percepción del tipo de situaciones que serían presuntamente "ideales", del papel que jugaría la "realidad" y de por qué la razonabilidad sería una expresión de la búsqueda de un equilibrio entre ambas.

Dado que los seres humanos somos, como todo organismo vivo, básicamente agentes racionales-egoístas (vale decir, agentes motivados fundamentalmente por el autointerés o deseo de maximizar el propio beneficio, y que a la vez contamos con habilidades cognitivas estratégicas implícitas o explícitas para alcanzarlo), una situación ideal sería en principio aquella de la cual podamos obtener el máximo provecho posible con el mínimo coste (por ejemplo, lo ideal sería pagar un precio ínfimo por un producto o servicio –o, en el límite, obtenerlo gratuitamente–; que nuestros reclamos telefónicos sean atendidos de inmediato sin hacernos esperar nada; recibir dinero prestado y devolverlo a larguísimo plazo sin ningún interés; etc.). Pero al mismo tiempo sabemos que nuestras aspiraciones a obtener el máximo beneficio están limitadas por los deseos e intereses de otros individuos con las mismas aspiraciones egoístas y el mismo derecho a satisfacerlas que nosotros. Sabemos, además, que no es posible obtener gratuitamente aquello que tiene un costo objetivo (tanto material como resultante de la suma de múltiples esfuerzos), y que los demás no estarán dispuestos a otorgarnos un beneficio gratuito, esto es, sin obtener nada a cambio. Básicamente, nuestra interacción con el ambiente social nos permitió abstraer reglas implícitas básicas de reciprocidad, no solo como una mera constatación de la existencia fáctica de tales reglas (sabemos por experiencia que en la

mayoría de los casos no es posible obtener ventajas gratuitas sin ofrecer ninguna contraprestación, del mismo modo que esperamos recibir una compensación por aquello que damos), sino también como un mandato normativo internalizado. En tal sentido, lo 'razonable' sería obtener una compensación justa por nuestras acciones (por ejemplo, es razonable que nos cobren por un producto o servicio y no seríamos razonables si pretendiéramos obtenerlo gratis, pero tampoco corresponde pagar un sobreprecio por el mismo –lo razonable sería, en todo caso, que el precio del producto o servicio sea proporcional a la calidad del mismo, al costo de los materiales y la mano de obra necesaria para obtenerlo, etc.–). En tal sentido, el criterio implícito de razonabilidad que sustentamos estaría ligado a un sentido de justicia (retributiva y distributiva), que a su vez se fundaría en un principio básico de reciprocidad: es razonable que todos aportemos una cuota similar de esfuerzo para alcanzar una meta común, que se nos retribuya de manera justa por nuestro trabajo, que dicha retribución sea proporcional, por ejemplo, a la inversión en tiempo y esfuerzo que dicho trabajo demanda, o a los años de estudio invertidos para poder realizarlo, etc. También sería razonable la distribución equitativa de un bien o recurso colectivo entre un grupo de sujetos si la inversión parcial de cada sujeto fue la misma (por ejemplo, si todos pagan impuestos en cantidades iguales, todos deberían acceder por igual a los beneficios resultantes), pero a su vez también sería razonable que los impuestos que paga cada sujeto sean proporcionales a sus ingresos etc. De todos estos ejemplos y muchos otros cabe desprender que en ciertas circunstancias pueden aplicarse criterios alternativos de razonabilidad igualmente atendibles, ya que son muchas las variables que es posible tomar en cuenta a la hora de buscar criterios asociados a reparticiones o compensaciones justas[4].

---

4 Actualmente me encuentro procesando los datos de una reciente investigación en filosofía experimental aplicada en una población de estudiantes secundarios en la que utilizo diversos dilemas morales sobre justicia distributiva cuyas alternativas de respuesta implican opciones a favor de diversos criterios (todos ellos razonables) de repartición de recursos (dar a todos por igual, priorizar al que menos tiene, priorizar al que realizó un mayor esfuerzo o tiene más talento, etc.).

Ahora bien, estos ejemplos ideados por Tobia parecen referirse indirectamente a las opciones que juzgamos razonables para nosotros mismos. ¿Podrían aplicarse también a los juicios morales sobre las acciones de otras personas?, ¿qué sucede cuando juzgamos acciones en 'tercera persona', vale decir, acciones que no afectan nuestros intereses? (por ejemplo, si un sujeto, que provocó un accidente involuntariamente, tuvo o no los cuidados razonables que hubieran podido prevenir o minimizar su ocurrencia). Mi posición al respecto es que, dada nuestra capacidad de juzgar acciones en calidad de 'espectadores imparciales' (en términos de Adam SMITH, 1937), estamos habilitados para, como diría Kant, transformar en norma universalmente aplicable aquellos imperativos ligados a un compromiso contractual tácito entre las partes (en este caso los miembros de cualquier sociedad humana), imperativos necesarios para que cada quien pueda defender sus propios intereses del abuso de los demás, y al mismo tiempo garantizarles a los otros la misma protección frente al potencial abuso que podamos cometer nosotros mismos. Para tal fin se vuelve imperioso que todos estemos, en virtud de un contrato social implícito, sujetos a ciertos deberes y derechos, y que dicho compromiso sea cumplido por igual por todos los miembros de la 'coalición' (término utilizado en teoría de juegos, aplicable metafóricamente a las sociedades humanas en general, en la medida en que supone que estas están conformadas por individuos que cooperan entre sí para lograr la consecución de objetivos que no podrían alcanzar individualmente por sus propios medios, para lo cual asumen ciertos compromisos normativos a condición de que los demás también los asuman).

En tal sentido, es posible asimilar al menos alguno de los criterios implícitos de razonabilidad propuestos y detectados empíricamente por Tobia a la hipótesis del altruismo recíproco de TRIVERS (1971), según la cual en nuestra especie (y en muchas otras especies sociales) habría evolucionado un tipo particular de altruismo o cooperación condicionado por la reciprocidad, vale decir, cuyo sostén y continuidad depende de que el comportamiento propio y ajeno se ciña a las mismas normas de reciprocidad colectiva. Solo de este modo el altruismo sería biológicamente 'redituable', vale decir, estaría al servicio del autointerés (propósito final al que tiende todo organismo vivo), ya que la inversión y las restricciones que cada miembro del grupo se autoimpone solo pueden tener sentido

bajo la expectativa de alguna compensación futura (la expectativa de que los demás también cooperen, y de este modo todos se beneficien del fruto de dicha cooperación). De este modo, lo 'razonable' tanto en términos empíricos (lo que *de hecho* suele hacer la mayoría de la gente, y al mismo tiempo lo que solemos esperar de los demás) como en términos prescriptivos (el comportamiento adecuado que *debemos* promover al interior de las sociedades a fin de que estos sistemas basados en la cooperación a gran escala se sostengan en el tiempo) sería alcanzar un equilibrio entre aquello que satisfaría idealmente nuestras expectativas (alcanzar los mayores beneficios posibles sin que ningún obstáculo se interponga en nuestro camino) y la realidad (los demás sujetos también aspiran a maximizar sus propios beneficios y para lograr su cometido pueden someternos a engaños, explotación, manipulación, fraude, dominación, etc.). De este modo, razonable sería simplemente aquel comportamiento, decisión, juicio de valor o razonamiento que busca alcanzar un equilibrio entre los intereses de las partes, asumiendo que en toda transacción ambas partes obtienen ganancias parciales y a su vez deben invertir o perder algo a cambio.

Cabe asumir que probablemente dicho criterio no agote por completo todas las posibles nociones de razonabilidad relevantes para el ámbito jurídico, ya que un comportamiento puede desviarse de parámetros 'razonables' no solo cuando el agente pretende obtener ventajas excesivas para sí mismo sin considerar la necesidad de alcanzar un equilibrio entre sus intereses y los de la contraparte. Por ejemplo, también puede considerarse no razonable o anormal algún tipo de comportamiento patológico o irracional. En este caso la falta de 'razonabilidad' estaría asociada a una suerte de déficit cognitivo que impide a las personas razonar o comportarse adecuadamente. Los casos analizados por Tobia sobre razonabilidad en contextos de negligencia (cuando se evalúa si un agente que provocó un daño accidentalmente había tomado o no medidas precautorias razonables para evitarlo) podrían interpretarse, dependiendo de la situación, en los dos sentidos esgrimidos. En algunos casos un acto de negligencia podría estar asociado indirectamente al egoísmo en el sentido de desidia o despreocupación por el prójimo (por ejemplo, una empresa que por ahorrar dinero no toma los recaudos de seguridad necesarios para proteger a sus operarios de posibles accidentes laborales). En situaciones como esta la negligencia estaría asociada a que el empresario antepuso sus propios

intereses egoístas, asociados al lucro, al interés por la salud y la seguridad de terceros. Pero en la mayoría de los casos la negligencia suele ser evaluada como un comportamiento imprudente por mera irracionalidad o 'estupidez', con lo cual la 'falta de razonabilidad' alude más bien a un déficit cognitivo y lo que se critica de dicho comportamiento no es tanto una falta moral como una suerte de 'desinteligencia', aun cuando la misma también suela ser juzgada moralmente (véase Zavadivker, 2015).

En síntesis, lo que quiero poner de manifiesto mediante estos ejemplos es que, además de la noción implícita de razonabilidad entendida como búsqueda del equilibrio entre el interés propio y los intereses ajenos, existiría otra acepción común del término, relevante también en contextos legales y jurídicos, que aludiría a la posesión de las capacidades cognitivas suficientes (y su uso estratégico eficaz) para un correcto desempeño en el entorno social (por ejemplo, para evitar desenlaces desafortunados que perjudiquen a otros). También suele considerarse en este contexto a la persona que no razona correctamente, o que no 'entra en razones', pero un análisis detallado de esta acepción requeriría tomar en cuenta muchos aspectos que exceden los límites de este trabajo.

Conclusiones

Este capítulo partió de las siguientes preguntas disparadoras: cuando los psicólogos morales abordan una investigación experimental, ¿se limitan a describir neutralmente el desempeño de los sujetos –en alguna tarea que involucra comportamientos, decisiones o valoraciones morales– o más bien *evalúan* dicho desempeño con arreglo a ciertos parámetros?, ¿dichos parámetros son a su vez morales, vale decir, implican la adhesión a alguna teoría moral implícita? La evaluación moral implícita en virtud de la cual los propios investigadores juzgan el desempeño de los sujetos experimentales, ¿coincide a su vez con las evaluaciones morales espontáneas que haría cualquier sujeto, vale decir, responden a un estándar normativo implícito propio de la arquitectura psicológica de nuestra especie?

Si tal fuera el caso es posible percibir la circularidad implicada en el proceso: cuando el investigador accede al dato empírico (desempeño moral de las personas) lo interpreta y evalúa valorativamente desde ciertos estándares implícitos, pero dichos estándares (criterios para juzgar acciones,

creencias y valoraciones morales) pueden a su vez ser 'descubiertos' o puestos de manifiesto en las propias investigaciones empíricas, si mediante estas se demuestra que los mismos parámetros que asume el investigador están también espontáneamente presentes en los juicios morales de la gente en general.

Por otra parte, y a pesar de haber intentado plasmar esta idea a lo largo del texto, cabe insistir en los distintos niveles en los que hemos intentado aplicar las nociones de *normalidad* y su asociación con el criterio de *razonabilidad*:

– En una primera instancia (llamémosle nivel *sustantivo*) aplicamos dichos conceptos a la evaluación que hace un sujeto de la razonabilidad del comportamiento, las creencias y los razonamientos *efectivos* de otros agentes. Podríamos ejemplificarlo mediante sentencias tales como 'Fulano actuó *razonablemente* cuando hizo *x*', 'En tales circunstancias era *razonable* que Sutano pensara o creyera *x*', o 'Mengano tuvo una reacción *razonable* frente a *x* sucesos'.

– En una segunda instancia se sitúan los juicios acerca del modo como un sujeto evalúa la manera de juzgar de otros (por ejemplo, cuando evaluamos la sentencia de un juez o de un jurado, vale decir, cuando consideramos las evaluaciones que estos hicieron de una situación). Por ejemplo, 'El juez opinó *razonablemente* que los elementos de prueba no eran suficientes para condenar al acusado', o 'El otorgamiento de libertad condicionada al preso *x* no fue una medida *razonable* teniendo en cuenta sus antecedentes criminales'.

– En una tercera instancia, nos hemos referido al juicio que hacen los propios psicólogos morales y filósofos experimentales cuando recaban evidencias empíricas sobre el comportamiento promedio de los sujetos e interpretan y evalúan esos datos a la luz de algún criterio (generalmente implícito) de 'normalidad'. En estos casos, aplicar el criterio de 'razonabilidad' al desempeño observado implicaría emitir un juicio explícito de valor que no sería lícito en el contexto de una investigación descriptiva. En su lugar, lo que los psicólogos morales suelen sostener explícita o implícitamente es que tales o cuales respuestas conductuales son 'normales' y tales otras se desvían de la norma. Por ejemplo, el título del artículo "Harming kin to save strangers: further evidence for abnormally utilitarian moral

jugdments after ventromedial prefrontal cortex"[5], hace referencia a juicios morales 'anormalmente utilitaristas' como consecuencia de una lesión cerebral). Si bien en cierto modo el criterio adoptado en el ejemplo (y al que aparentemente suelen apelar los psicólogos experimentales) parece ser la mera comparación estadística de los datos obtenidos con el desempeño de la media poblacional (vale decir, si los sujetos evaluados actúan simplemente como lo haría el sujeto promedio, o bien se desvían de la curva normal), la realidad es que, más allá de las frecuencias estadísticas, a todos nos 'hacen ruido' los juicios excesivamente utilitaristas. El ejemplo del artículo hace referencia a que los sujetos con daños en la corteza prefrontal ventromedial eligen la respuesta utilitarista en el famoso 'dilema del puente' (lanzar al señor obeso a las vías del tren desde un puente para detener su paso y así salvar la vida de cinco personas a punto de ser arrolladas). Dicha opción produce en la amplia mayoría de las personas una aversión emocional inmediata que insta a su automático rechazo y a la percepción de su carácter inmoral, ya que la abstención de causar daño directo a un sujeto inocente se impone, por su peso emocional, con mayor fuerza que el cálculo instrumental, fruto de un razonamiento más lento, de que tal acción acarrearía un valor agregado. Ahora bien, dicho criterio de anormalidad asociado a la incorrección moral, al tratarse de un sesgo propio de la especie, sin duda también afecta al investigador, quien cree *espontáneamente* que tal juicio es incorrecto, y no solo por el mero hecho de que se desvía de la respuesta estadísticamente mayoritaria. Lo cree *a priori.*

En síntesis, los criterios de normalidad y razonabilidad operan en planos distintos que muchas veces se confunden. El propio Tobia no establece distinciones claras entre estos planos, generando una confusión entre lo que la gente considera 'acciones reales, ideales y razonables'; la realidad, idealidad y razonabilidad de los propios juicios de la gente sobre esos tópicos; y, a su vez, los mismos criterios aplicados al investigador que evalúa los juicios de los participantes. Pero la razón de la confusión estriba precisamente en que todos ellos son *interdependientes* y parecen formar parte de una misma lógica: si una acción es considerada razonable (o 'normal'),

---

5 Título de un artículo de B. C. Thomas, K. E. Croft y D. Tranel publicado en la revista *Journal of Cognitive Neuroscience* en 2011.

también sería razonable (y 'normal') juzgar que dicha acción lo es, y a su vez sería razonable que un psicólogo moral evalúe la razonabilidad del sujeto experimental cuando este evaluó razonablemente un comportamiento razonable. La razón de fondo es simple: todos los agentes implicados en el proceso evaluativo (el agente que actúa, el que evalúa y el que evalúa al que evalúa) estamos 'cortados por la misma tijera', ya que somos ejemplares de la misma especie y, por ende, poseemos una arquitectura cognitiva similar.

Otro aspecto abordado en este capítulo es el de las implicancias para el ámbito legal y jurídico de la insalvable circularidad entre descripción y normatividad, y en qué medida el criterio de 'razonabilidad', entendido como un término medio entre un ideal normativo y los datos de la realidad, puede arrojar luz y constituir la mejor guía para orientar las decisiones en esos terrenos. En la reflexión que sigue procuraré demostrar esta afirmación.

Nuestros sistemas legales y jurídicos reflejan a las claras la necesidad de encontrar el equilibrio entre un conjunto de pautas normativas a priori, abstractas y universalmente aplicables, y las evidencias empíricas que aportan datos concretos, por ejemplo, sobre la potencial peligrosidad de ciertos individuos por su pertenencia a ciertos grupos (por poseer ciertas patologías de base, por su género, por su condición social, por su formación ideológica, etc.). El sistema jurídico está pensado para proteger a cualquier individuo, más allá de sus características particulares, de juicios previos carentes de pruebas suficientes y basados por ende en sesgos, prejuicios o etiquetamientos –por lo general ligados a la pertenencia del acusado a cierto grupo–, pero al mismo tiempo sabemos que ese exceso de neutralidad e imparcialidad, indudablemente correcto desde el punto de vista del *deber ser*, puede desencadenar en los hechos resultados peores de los que se hubieran obtenido en caso de haberse tomado medidas preventivas basadas en la posible predicción del comportamiento de tales sujetos. Dichas medidas, al no basarse en comportamientos efectivos sino en la anticipación del posible desempeño futuro sobre la base de ciertos indicios, sin duda serían 'prejuiciosas', pero se basarían a su vez, o bien en datos empíricos (por ejemplo, en el porcentaje de reincidencia de los miembros de ciertos grupos, como violadores, femicidas, pederastas, asesinos seriales, etc.), o bien en teorías científicas avaladas a su vez por evidencia empírica (por ejemplo, en patologías de base que impedirían a algunos sujetos su reinserción social). Un ejemplo reciente muy claro de modificaciones en

la legislación y los códigos procesales basadas en evidencias fácticas es la introducción de la perspectiva de género. Pese a su ropaje progresista, la legislación con perspectiva de género no es más que la introducción de un prejuicio de género en el seno del sistema jurídico (en principio universal y abstracto, en la medida en que fue pensado para brindar, al menos formalmente, las mismas garantías a cualquier persona por igual). Dicho prejuicio, que ha llevado en los hechos a la negación, con fines preventivos, de ciertos derechos y garantías –en particular, la presunción de inocencia– al hombre sospechado de violencia física o abuso sexual, se fundaría en la consideración de que las evidencias fácticas indican a las claras una marcada tendencia de los hombres a ejercer violencia física y sexual contra las mujeres, evidencias que a su vez estarían respaldadas por teorías que hacen referencia, o bien a la propia naturaleza biológica masculina, o bien –con una incidencia claramente mayor en el discurso feminista hegemónico actual– a la dominación histórica del hombre sobre la mujer como resultado de una educación machista y patriarcal. La abrumadora cantidad de datos empíricos sobre experiencias desafortunadas de mujeres que no fueron escuchadas o cuyas denuncias por violencia de género no generaron medidas por falta de pruebas y luego terminaron siendo víctimas de femicidio, y el masivo movimiento de mujeres bajo consignas como "*Me Too*" o "*Yo sí te creo*", instaron al sistema judicial a modificar su criterio y considerar el testimonio de una potencial víctima como suficiente para tomar medidas judiciales en contra de los acusados, reemplazando el criterio de presunción de inocencia por el de 'presunción de culpabilidad' (todo hombre acusado de violencia de género es culpable a menos que demuestre lo contrario) e invirtiendo así la carga de la prueba (es el acusado quien debe proporcionar las pruebas que demuestren su inocencia, en lugar de que sea el demandante quien presente las pruebas que demuestren la culpabilidad del acusado). Lamentablemente, en la actualidad, en países en los que la perspectiva de género modificó drásticamente el sistema judicial otorgando, no los mismos derechos, sino privilegios desmesurados a las mujeres –y al mismo tiempo privando de derechos y garantías mínimas a los hombres–, están saliendo a la luz cada vez más evidencias sobre falsas denuncias de acoso sexual y violencia de género, por razones tales como el mero oportunismo –por ejemplo, mujeres que buscan ventajas en un juicio de divorcio, como la tenencia de los hijos o beneficios en la repartición de

bienes–, por despecho –un modo de vengarse de otros presuntos daños como la infidelidad–, etc. Se conocen muchos casos, verbigracia, de madres que obligan a sus hijos, bajo amenazas y maltrato, a declarar en contra de sus padres acusándolos de abuso sexual o maltrato físico. En el futuro la acumulación de evidencias sobre denuncias falsas probablemente inste a la legislación y al sistema judicial a volver a adaptar sus leyes y su jurisprudencia a estos nuevos datos de la realidad (que por lo pronto indicarían que la condición de hombre no implica necesariamente la de victimario, ni la condición de mujer la de víctima). En síntesis, los ámbitos legales y judiciales se debaten permanentemente entre un conjunto de prescripciones abstractas que en principio buscan resguardar a todo sujeto por igual de las posibles arbitrariedades del sistema, y los datos empíricos que indican que, en algunos casos, prejuzgar a ciertos individuos por su pertenencia a un grupo puede dar en la práctica mejores resultados que abstenerse de emitir juicios previos; pero al mismo tiempo dichos prejuicios siguen acarreando el peligro de hacer 'pagar a justos por pecadores'.

En este contexto, el criterio de razonabilidad (que históricamente en la literatura jurídica apareció bajo conceptos análogos tales como 'sensatez' o 'prudencia') sería la expresión más adecuada del intento de encontrar el punto medio entre estándares normativos abstractos que solo serían aplicables en condiciones ideales (en el ejemplo citado, el principio rector sería el de 'imparcialidad'); y, no solo los datos de la realidad empírica, sino también el anhelo pragmático y adaptativo de acomodarnos a dicha realidad para obtener mejores resultados en la práctica que los que obtendríamos si solo rigiéramos nuestro comportamiento por prescripciones abstractas e inamovibles. Ahora bien, como lo demuestra el ejemplo, lo más sensato parece ser entender por 'datos empíricos' los aspectos particulares y específicos de cada caso por separado, sin dejarnos llevar por burdas generalizaciones basadas en prejuicios –muchas veces avalados por teorías omniexplicativas– que lleven a un individuo a ser juzgado por su grupo de pertenencia (sobre todo cuando ese 'grupo' es lo suficientemente amplio como para poder englobar, en principio, a cualquier clase de individuo, como es claramente el caso de los 'hombres' y las 'mujeres').

Analizaremos brevemente este nuevo criterio de razonabilidad, a medio camino entre el juicio completamente imparcial y libre de prejuicios[6] y los datos empíricos (incluyendo experiencias propias y ajenas) que nos indican que cierta dosis de prejuicio y discriminación (vale decir, de desconfianza en ciertos grupos) posee un claro valor pragmático y adaptativo al ayudarnos a prevenir potenciales daños futuros. Dicho criterio es también aplicable a los tres niveles de evaluación del agente moral analizados:

– Sería razonable que las personas comunes exhiban el trato más igualitario posible con los demás sin guiarse por las apariencias o los prejuicios sociales, pero también que mantengan un grado prudencial de cautela sobre la base de la información social ligada a la potencial peligrosidad de ciertos grupos, lugares, etc.[7].

---

6 Este tipo de juicio se plasmaría en la tan extendida expresión actual 'políticamente correcto'. Sin embargo, si bien el criterio de corrección política alude a la eliminación de todo tipo de prejuicio o sesgo discriminatorio en contra de un grupo en particular, la realidad es que los criterios de corrección política suelen variar históricamente en función de cuál sea el discurso dominante vigente por lo general en los círculos intelectuales y progresistas. En otras palabras, quien suele fijar los parámetros de 'corrección política' es el discurso oficial de los sectores representativos del pensamiento de 'izquierda' en un determinado momento histórico y cultural, aplicándolos no a todos los grupos por igual, sino solo a aquellos grupos percibidos como 'minorías oprimidas y excluidas'. Por ejemplo, actualmente la corrección política no parece aplicarse a la hora de juzgar a los hombres blancos occidentales y heterosexuales, ya que se considera que estos forman parte del grupo hegemónico y opresor y por ende merecerían ser objeto de condena social y trato discriminatorio.

7 Obviamente, en los juicios morales del hombre común la balanza se inclina mucho más del lado del prejuicio, la estigmatización y el etiquetamiento social que del lado de la imparcialidad y la neutralidad (otro argumento a favor de que los parámetros de corrección moral no pueden depender de lo que la mayoría estadística opina). Incluso, así como el 'hombre común' suele reproducir los prejuicios dominantes más comunes en su sociedad (reflejando espontáneamente el pensamiento conservador o de 'derecha'), también los sectores más instruidos pero fuertemente condicionados por prejuicios teóricos y políticos (el hombre de 'izquierda', por decirlo brevemente, ya que no es posible acá entrar en detalles) suelen aplicar estigmas y etiquetamientos implacables a quienes sustentan ideologías contrarias a las propias, o a quienes perciben como miembros de grupos dominantes, hegemónicos y opresores.

– También sería razonable, como ya lo analizamos, que las leyes, la jurisprudencia, las decisiones políticas, los fallos judiciales, etc. respeten los principios básicos de imparcialidad e igualdad ante la ley –que incluyen el derecho de todo ciudadano a ser considerado inocente hasta que se demuestre lo contrario–; pero también sería 'sensato' tomar recaudos preventivos contra sujetos pertenecientes a ciertos grupos 'problemáticos' (siempre y cuando el etiquetamiento dentro de esos grupos no dependa, por ejemplo, de prejuicios sociales, raciales, ideológicos o de género, sino de aspectos más puntuales tales como la posesión de determinada patología), a los fines de resguardar la seguridad futura de algunas víctimas potenciales.

– ¿Y cómo entender en el nivel metaético (el que sustentan los psicólogos y filósofos morales experimentales) esta misma búsqueda de un equilibrio entre juicios puramente racionales y juicios funcionales desde el punto de vista pragmático y adaptativo? En la primera parte de este escrito hemos mencionado entre las teorías prescriptivistas aquella que concibe al agente moral ideal como un sujeto puramente racional guiado, para decirlo en términos muy generales, por el imperativo kantiano. Dicha racionalidad se aplicaría tanto en el terreno del comportamiento moral ('obrar conforme a la razón y contra inclinación') como en el juicio y el razonamiento moral, que deberían estar inspirados en principios racionales tales como los de justicia, imparcialidad, reciprocidad, etc. y evitar cualquier tipo de falacia o sesgo distorsionante capaz de inclinar la balanza, por ejemplo, a favor de nuestros propios intereses (personales, de clase, de género, étnicos, etc.), como así también evitar las distorsiones provocadas por emociones como el miedo, la ira, el asco, etc. Existen muchas investigaciones en psicología y filosofía experimental que buscan precisamente detectar todos esos sesgos en el comportamiento y el razonamiento humanos con el propósito de hacerlos explícitos a los fines de que puedan someterse a un control reflexivo y consciente, lo que permitiría, por ejemplo, evitar que distorsionen las decisiones de los jueces o del jurado en un fallo; detectarlos en los discursos políticos, etc. Para esta perspectiva, entonces, los sesgos en el razonamiento moral espontáneo de las personas serían simplemente 'errores cognitivos', en algunos casos inspirados incluso en propósitos inmorales inconscientes, y lo deseable desde el punto de vista normativo sería procurar tornarlos conscientes para fines de eliminarlos. Pero existe también una perspectiva metaética alternativa de corte naturalista (y más

específicamente evolucionista), que, si bien en principio aparenta ser meramente descriptiva (apuntando además a la explicación causal del fenómeno de la moralidad como una estrategia más de supervivencia, específica de nuestra especie), en el fondo asimila lo 'bueno' (no en sentido moral, sino en el sentido de provechoso o deseable) a lo que resulta adaptativo o biológicamente eficaz, vale decir, aquello que promueve la supervivencia y el éxito reproductivo en un contexto dado. Dicho contexto presenta determinados desafíos y presiones selectivas, y nuestra arquitectura cerebral (incluyendo nuestros comportamientos, preferencias, creencias, juicios de valor y razonamientos morales) estaría biológicamente programada para responder eficazmente a ellas. Desde esta perspectiva podría considerarse que todo lo 'dado' (tales comportamientos y valoraciones espontáneas del común de la gente) existe como tal porque tiene o tuvo en el pasado algún valor adaptativo, de modo que lo 'natural' debe ser interpretado a priori como 'bueno'. Para quienes sustentan esta posición metaética, los sesgos en el razonamiento o los juicios morales estarían allí por poseer algún valor adaptativo intrínseco, vale decir, por su capacidad para responder eficazmente a los desafíos que emanan del ambiente social, que es nuestro nicho ecológico por excelencia. En tal sentido, y retomando las nociones sobre altruismo recíproco de Trivers anteriormente expuestas, desde una perspectiva evolucionista podría concebirse al criterio de razonabilidad como una respuesta adaptativa óptima en la medida en que considera al mismo tiempo el deber-ser (las condiciones necesarias de autolimitación del comportamiento egoísta bajo las cuales nuestra especie puede subsistir del modo más eficaz posible en un entorno social conformado por agentes esencialmente motivados por el autointerés) y la realidad (el grado de resguardo necesario contra la potencial amenaza del prójimo, o bien las limitaciones reales a nuestros anhelos de obtener el mayor provecho posible de las situaciones). Nótese que en este caso estoy utilizando la noción de deber-ser o situación moralmente ideal en un sentido diferente –y más precisamente contrario– al que utilicé al analizar las implicancias del experimento de Tobia a la luz de la psicología evolucionista y la teoría de juegos: en el caso anterior la situación ideal era la que reportaría los mayores beneficios egoístas al evaluador; mientras que acá entiendo por situación ideal aquella en la cual el comportamiento de todos los agentes morales es el óptimo desde el punto de vista ético (vale decir, cuando es

posible esperar con plena confianza que todos los demás sujetos de una coalición respetarán todas las normas de reciprocidad colectiva, y, por ende, no necesitamos ponernos al resguardo de potenciales tramposos [*cheaters*], abusadores, etc.). A pesar de la aparente contradicción, en ambos casos estamos hablando de predisposiciones adaptativas, ya que para cualquier sujeto es adaptativo obtener las mayores ventajas posibles de la vida en sociedad, pero al mismo tiempo sabe que las consecuencias de no acatar normas colectivas pueden no ser adaptativas, y por otra parte, advierte (aunque sea inconscientemente) la conveniencia de que todos los miembros del grupo –incluyéndolo– acaten dichas normas a fin de generar un ambiente de confianza mutua que torne posibles y predecibles las interacciones sociales que redundarían en beneficios para todos, y se resguarde lo mejor posible de la potencial amenaza de los demás.

REFERENCIAS

BAGEHOT, W., 1873: *The English Constitution.* Oxford University Press, 2009.

GREENE J., NYSTROM, L., ENGELL, A. DARLEY, J. COHEN, J., 2004: "The neural basis of cognitive conflict and control in moral judgment", *Neuron,* 44: 389-400.

KNOBE, J. y NICHOLS, S., 2008: "An experimental philosophy manifiesto", en KNOBE, J. y NICHOLS, S. (eds.), *Experimental Philosophy.* Vol. 1. New York: Oxford University Press.

SMITH, A., 1937: *A Theory of Moral Sentiments.* New York: Modern Library.

THOMAS, B. C., CROFT, K. E. y TRANEL, D., 2011: "Harming kin to save strangers: further evidence for abnormally utilitarian moral jugdments after ventromedial prefrontal cortex", *Journal of Cognitive Neuroscience,* 23 (9): 2186-2196.

TOVIA, K., 2018: "How people judge what is reasonable", *Alabama Law Review*, 70 (2): 293-359.

TRIVERS, R. L., 1971: "The evolution of reciprocal altruism", *The Quarterly Review of Biology*, 46 (1): 35-57.

Zalesne, D., 1997, "The intersection of socioeconomic class and gender in hostile housing environment claims under Title VIII: Who is the Reasonable Person?", *B.C. L. Rev.*, 38 (15): 861-863.

Zavadivker, M. N., 2014: *Homo eticus. Las bases biológicas del comportamiento prosocial*, S. M. de Tucumán: La Monteagudo.

Zavadivker, M. N., 2015: "'No es lo que conoces, sino lo que ponderas'. Una propuesta alternativa sobre las bases neuronales y cognitivas implicadas en la 'suerte moral'", *Panamerican Journal of Neuropshychology*, 9 (1): 41-67.

IX.
¿Es el sometimiento de la víctima una alternativa al odio como estado mental?

MARÍA LAURA MANRIQUE*

Sin lugar a dudas, las conductas desencadenadas por el odio nos provocan un intenso y especial rechazo. Por ejemplo, en una conocida tragedia, uno de sus personajes centrales, Yago, movido por su envidia y odio hacia Otelo, elabora una compleja trama que resulta en el asesinato de Desdémona. En esa clásica obra de Shakespeare, Otelo es el autor de ese homicidio, pero la manipulación y el engaño de Yago provocan un repudio tan intenso como (o aún más fuerte que) el crimen de Otelo.

El odio, al igual que los celos, la envidia, la compasión, etc., es un *motivo emocional*, es decir, una emoción que mueve a la acción y, por esa razón, tiene un papel importante en la atribución de responsabilidad penal. Así, en las últimas décadas, el odio ha sido incorporado en numerosas legislaciones penales, tanto en la recepción de ciertos tipos delictivos específicos (*i.e.*, los llamados 'delitos de odio') como también en la incorporación del odio como una agravante específica de otros delitos. Por ejemplo, el artículo 80, inciso 4, del código penal argentino establece que se impondrá prisión perpetua a quien matare a otro por "placer, codicia, odio racial, religioso, de género o a la orientación sexual, identidad de género o su expresión".

La complejidad del odio como motivo emocional genera importantes problemas de diferente índole. Por ejemplo, la incorporación del odio en la legislación inmediatamente suscita interrogantes de tipo probatorio, por ejemplo, ¿qué tipo de eventos o hechos es necesario acreditar para probar que una conducta fue realizada por odio? Otros interrogantes, podría señalarse con cierta cautela, pertenecen al ámbito de la dogmática penal. En particular, es posible subrayar dos grupos de problemas específicos. Por una parte, el odio comparte con otras emociones una naturaleza de variable intensidad, que en ocasiones se presenta como una fuerza irresistible, que

---

* Magíster en Derecho Penal y Ciencias Penales de la Universidad Pompeu Fabra (Barcelona, España). Doctora en Derecho de la Universidad Pompeu Fabra. Fue becaria posdoctoral del Instituto de Investigaciones Filosóficas de la Universidad Nacional Autónoma de México – UNAM (México) y de la Universidad de Girona (España). Actualmente es investigadora adjunta del Consejo Nacional de Investigaciones Científicas y Técnicas – CONICET (Argentina).

va 'más allá' del control del agente. Por otra parte, frecuentemente el odio se articula como una razón para actuar y nos permite comprender la conducta del agente a la luz de ciertos motivos específicos. Ambas cuestiones son problemáticas en un sistema de derecho penal que pretende limitarse a castigar únicamente el resultado disvalioso de las *acciones* de los agentes, ya que el primero de ellos apunta a situaciones que escapan al control de los individuos, mientras que el segundo parece presuponer que el odio añade un *plus* al injusto realizado.

A lo largo de este trabajo no discutiré la cuestión general de si está justificado considerar a las motivaciones de un agente como un agravante de su responsabilidad penal[1]. Tampoco me detendré en la cuestión más específica de qué motivos emocionales son relevantes para agravar o atenuar la responsabilidad por ciertos delitos. De este modo, dejaré de lado las cuestiones relativas a la justificación de la incorporación de ciertas agravantes. Así, en un sistema penal que califica de manera más grave al odio racial en casos de homicidio, pero no agrava ese delito cuando es cometido por celos, el asesinato de Desdémona a manos de Otelo merecería menor reproche del que hubiese recibido Yago en caso de que hubiese estrangulado a Otelo, movido por la envidia y el odio racial. Aunque esta

---

1 No entraré a discutir la legitimidad de la prohibición de la incitación al odio como la establecida, por ejemplo, en el artículo 510 del código penal español. Según el mismo, serán reprimidos "Quienes públicamente fomenten, promuevan o inciten directa o indirectamente al odio". Creo que este tipo de regulación genera problemas diferentes y, en ocasiones, las dificultades de los tipos penales están en la amplitud y variedad de la descripción de las acciones prohibidas más que en los motivos en particular.
Tampoco me ocuparé aquí de otros temas vinculados a la discusión de 'delitos de odio' como el discurso de odio, cuyo problema central es el límite a la libertad de expresión, ni de los llamados '*henious crimes*', que son aquellos delitos que son considerados horripilantes u odiosos por la comunidad independientemente del motivo por el cual fueron cometidos. Tampoco me ocuparé del modo en que se generan los prejuicios o estereotipos que deshumanizan a las víctimas. La presencia de estos puede ser central para explicar la dinámica de los delitos de odio, pero ellos se centran en aspectos diferentes a lo que me interesa resolver aquí. Para un desarrollo de estos aspectos véase, por ejemplo, Díaz López, 2013, o los trabajos compilados en Miró Linares (dir.), 2017.

consecuencia puede parecer contraintuitiva, solo será abordada de manera tangencial en esta investigación.

Limitaré mi trabajo al problema del odio como motivo emocional, y mi propósito es revisar una justificación específica de la agravante de odio, que desplaza la pregunta acerca de la relevancia de motivos del agente hacia consideraciones referidas a la víctima. A efectos de acotar el recorrido bibliográfico de este ensayo centraré mi análisis en una de las más importante propuestas contemporáneas, elaborada por el catedrático de derecho penal de la Universidad de Córdoba (Argentina), José Milton Peralta (2011, 2013)[2]. Esta propuesta pretende justificar la incorporación de la agravante de odio como motivo emocional pero, al mismo tiempo, admite que si esa justificación se limitase al hecho de que el odio es un motivo peculiar (*i.e.*, algo que vuelve al crimen, precisamente, 'más odioso'), entonces esa agravante sería una suerte de reproche por las 'malas intenciones' del autor. Por ello, Peralta defiende un enfoque que asocia el odio con el sometimiento de la víctima.

Sin embargo, intentaré mostrar que, a diferencia de lo que cree Peralta, la comprensión de los motivos como estados mentales no solo es más adecuada a la tesis que propone (el odio como una forma de sometimiento) sino que también se hace inescapable para su propia posición.

La estructura de este trabajo es la siguiente: en la sección 1, reconstruiré brevemente las ideas centrales del trabajo de Peralta; en la sección 2, haré algunos comentarios críticos a la idea de que quien castiga por el odio como estado mental está únicamente reprochando las malas intenciones de los agentes. Así mismo, mostraré los problemas que presenta entender el odio como sometimiento. Mi conclusión, brevemente expuesta, será la siguiente: una propuesta como la de Peralta (*i.e.*, el odio como sometimiento) es insuficiente en un nivel explicativo ya que no da cuenta de casos paradigmáticos de odio. A su vez, a nivel justificativo es inadecuada ya que depende, en última instancia, del reconocimiento de la relevancia de los motivos emocionales como estados mentales.

---

2 Acerca del impacto de la propuesta de Peralta, véanse, por ejemplo, los siguientes trabajos que recogen sus ideas centrales: Mañalich, 2016; Wilenmann, 2016; Pérez Manzano, 2018.

Finalmente, debe resultar claro que no pretendo abordar todos los argumentos disponibles en la importante literatura especializada sobre los delitos de odio que con argumentos retributivos o prevencionistas pretenden fundamentar el mayor reproche en los delitos de odio[3]. Aquí examinaré con algún detalle una estrategia de fundamentación de los delitos de odio distinta a la fundamentación con base en los estados mentales del agente y centrada en la pretensión de someter a su víctima que posee el agente[4].

1. LA TESIS DEL SOMETIMIENTO

En este apartado reconstruiré brevemente una tesis sobre la manera de entender los motivos en los códigos penales: *la tesis del sometimiento*. Quienes defienden esta tesis pretenden mostrar que agravar los delitos cometidos por odio está justificado, pero su justificación no se encuentra en los motivos como estados mentales de los agentes sino que el motivo, bien mirado, indica que es el sometimiento aquello que el autor pretendía obtener de la víctima al momento de cometer su delito. En otras palabras, este enfoque no coloca el acento en si el reproche de los motivos es, en última instancia, solo una manera especial de censurar los 'pensamientos malvados' de los agentes (*guilty mind*), sino en el sometimiento y vulnerabilidad de la víctima.

---

3 Para una reconstrucción de este tipo de argumentos véanse, por ejemplo, Díaz López, 2013; Peralta, 2012.

4 Existe una profunda discusión previa a los problemas que resalto en este trabajo. Se trata de si las motivaciones en general, deseos, intenciones, aspiraciones, etc., deben ser consideradas relevantes en general para establecer el reproche por ciertas acciones. Esta es una discusión que puede leerse en clave de la oposición entre teorías deontológicas y teorías consecuencialistas. Aunque, en general, la adopción de una tesis consecuencialista radical impediría tomar en cuenta ciertos elementos intencionales, se encuentran teorías de base consecuencialista que pretenden fundamentar el reproche de los motivos en, por ejemplo, el mayor daño que generaría la acción realizadas por odio. Para esta discusión más general y previa, véanse, por ejemplo, Tadros, 2013: 206-227; Díaz López, 2013: 135-240. Para una fundamentación de los motivos desde las distintas teorías de la pena, véase Peralta, 2012: 37-118.

Quienes defienden la tesis del sometimiento pretenden mostrar que castigar de modo agravado ciertos tipos de delitos (*e.g.*, el homicidio por odio racial, religioso o femicidio) está justificado en un derecho penal de acto pero su justificación no se encuentra, como tradicionalmente se había considerado, en los motivos como estado mental del agente. En verdad, los 'motivos' son penalmente relevantes solo en otro sentido o dimensión. Por ejemplo, cuando indican que el autor pretende someter a su víctima para que viva del modo que a él le parece adecuado. Precisamente, este rasgo, que no se encuentra en los homicidios comunes, justificaría el mayor reproche. Así, en palabras de Peralta:

> La manera en que las víctimas pueden evitar la agresión en estos casos es sometiéndose a la voluntad de un autor que quiere imponerles un modo de vida; la contracara es que el autor las mata porque no se han sometido (Peralta, 2013: 4).

La propuesta de Peralta se articula en torno al desarrollo de tres tesis independientes (ibíd.: 1-4):

1) *Tesis de la relevancia del odio*: los delitos de odio deben agravarse;

2) *Tesis negativa sobre los motivos como estados mentales*: castigar más severamente por los motivos que el agente posee –entendidos como estados mentales– es reprochar los pensamientos y, por lo tanto, no está justificado en un Estado liberal;

3) *Tesis positiva*: la noción de sometimiento explica correctamente por qué agravamos los delitos de odio y justifica su mayor reproche.

Peralta divide la fundamentación de estos delitos en dos partes. La primera, presentada como una idea prospectiva, es donde cobra relevancia la noción de sometimiento y muestra una característica definitoria de los delitos de odio. La segunda, presentada como una idea retrospectiva, gira en torno a la falta de provocación. La noción de provocación es utilizada para mostrar qué diferencia los homicidios comunes de los delitos agravados.

1.1. Aspecto prospectivo

Peralta, en lugar de proveer de una definición sobre delitos de odio, comienza caracterizando el homicidio por odio a la identidad de género o

su expresión. Este es un caso típico de delito de odio y está tipificado en Argentina en el artículo 80, inciso 4, del código penal. Peralta sostiene que el autor basa su homicidio en que la víctima realizó alguna expresión de su identidad, *e.g.*, travestirse. Dicho de otro modo, la víctima *hace algo* que, desde el punto de vista del agente, no *debería* haber realizado (Peralta, 2013: 11).

> … lo que hace el autor con el homicidio por odio es, además de matar, censurar la acción de la víctima, manifestar que no existe derecho a travestirse, o por lo menos, que esa no es, mientras el autor exista, una opción inteligente (ibíd.: 11).

La cuestión es que, para evitar la muerte, la única opción que le queda estratégicamente a la víctima es llevar una vida donde no expresa su identidad de género, o vivir del modo en que el autor desea. La víctima debe *someterse* a los deseos del autor si pretende seguir viviendo. Así, la víctima pierde toda su autonomía para ejercer sus derechos. Peralta hace el mismo ejercicio con homicidios por odio religioso, racial o femicidio[5]. En todos esos casos, la idea es la misma: para dejar de ser víctima y no 'provocar' al agresor, la mujer, la persona de color o el musulmán, según el caso, deben someterse al modo de vida que al autor le parezca correcto que una persona de sus características debe llevar. Así, las víctimas no pueden ejercer su autonomía (ibíd.: 14).

En conclusión, la razón que explica y justifica el mayor reproche en los delitos por odio es el sometimiento que pretende ejercer el autor respecto de la víctima del homicidio (ibíd.: 4). Esta razón explica por qué en los homicidios simples no aparece la idea de sometimiento, *i.e.*, no hay un autor que pretenda que la víctima se subyugue a su modo de vida.

---

5 Peralta hace el mismo ejercicio con otro tipo de homicidios por odio (odio religioso, odio racial y femicidio); para lo que se pretende mostrar aquí creo que es suficiente con este ejemplo. Para el desarrollo de las ideas, véase Peralta, 2013: 11–13.

1.2. ASPECTO RETROSPECTIVO

¿Qué justifica el mayor reproche de las conductas homicidas realizadas por odio respecto de aquellos homicidios comunes si externamente el resultado es el mismo, *i.e.*, ambas víctimas terminan muertas? La respuesta de Peralta es que esta pretensión de justificación se torna más compleja desde el punto de vista retrospectivo e individual, pero la idea de sometimiento puede explicar la mayor gravedad de los homicidios por odio. En su opinión, la justificación de la agravante reside en que, más allá del injusto del resultado, la razón de la muerte de la víctima fue la reacción provocada en su agresor por el ejercicio de un derecho individual. Según Peralta:

> No se agravan los delitos por la inmoralidad del motivo como tal, sino que el motivo indica *otra cosa*, íntimamente vinculada, que sí da cuenta de la mayor gravedad del *hecho.* El motivo, en realidad, señala una idea de sometimiento (PERALTA, 2013: 14).

De este modo, la interacción anterior al hecho delictivo entre la víctima y el agente se convierte en un dato central para el análisis (ibíd.: 14). Siguiendo esa idea, Peralta utiliza las herramientas planteadas por la victimodogmática, en donde cobra especial relevancia la noción de *provocación.* Sostiene que en los homicidios simples la víctima ha realizado alguna conducta que puede entenderse como provocación (no en sentido estricto) o que genera un cierto conflicto. Si bien la víctima no ha perdido ningún derecho por esta provocación, existiría un conflicto previo que explica y hace inteligible el comportamiento del autor[6].

> … en todo homicidio simple existe alguna especie de provocación penalmente relevante y eso es lo que lo diferencia del homicidio calificado o asesinato donde no existe provocación alguna o no existe provocación relevante […] Los homicidios no ocurren sin motivos, sino que suelen ser la consecuencia de un conflicto (ibíd.: 17-18).

---

6 Esta idea es tomada por Peralta de Müssig. Véase PERALTA, 2013: 17; PERALTA, 2012: 202-212.

Así, los casos de homicidio simple son los casos en que existe un conflicto previo como, por ejemplo, el de quien mata a otro porque lo insultó al realizar una maniobra brusca conduciendo. En ausencia de una provocación existirá un homicidio agravado. En el homicidio por odio, la víctima realiza una conducta lícita como el ejercicio de un derecho individual que no puede entenderse como una provocación en sentido relevante, independientemente de la percepción que tenga el autor (por ejemplo, el autor se siente provocado porque una persona de color accede a un cierto cargo) (PERALTA, 2012: 215 ss.).

Para Peralta, los motivos expuestos en las agravantes del homicidio (art. 80 c.p. argentino) responden a la pregunta de por qué el autor actuó de esa manera. Su respuesta es que el autor debe ser reprochado más gravemente porque su conducta de ninguna manera puede atenuar el delito cometido. En síntesis, los homicidios son agravados porque existe esta ausencia de cierta responsabilidad de la víctima por la conducta del agente. Los delitos de odio como sometimiento son una especie de delitos agravados en donde el autor pretende que la víctima viva o se comporte de determinada manera y conforme a sus reglas. La agravante general tiene la utilidad, según Peralta, de explicar tanto por qué se agravan delitos que no son de odio como también homicidios por odio que no son de sometimiento (PERALTA, 2013: 21–23).

2. PROBLEMAS DE LA TESIS DEL SOMETIMIENTO

2.1. LOS MOTIVOS COMO MEROS PENSAMIENTOS

La tesis negativa sostenida por Peralta presupone que reprochar por motivos no está justificado porque los motivos –por así decirlo– no 'se reflejan en el mundo'. Ellos serían solo pensamientos, y por ello, vincularlos a la responsabilidad penal sería un modo de castigar meros estados mentales. En este sentido, es frecuente encontrar diferentes formulaciones de esta idea en la dogmática penal contemporánea[7], pero continuando con Peralta, él sostiene:

---

7 Para otro ejemplo de este tipo de argumentos, véase NINO, 1980: 153 ss.; HURD y MOORE, 2004: 1123.

> Una de las características de un Derecho penal de acto es, justamente, que impide que los pensamientos *qua pensamientos* puedan ser objeto de castigo estatal. La razón de esta restricción no radica en la calidad de esos pensamientos, que bien pueden parecernos más o menos elogiables o censurables, sino en que no es competencia del Estado evaluar aspectos de la vida de las personas que no trasciendan en un daño para terceros (PERALTA, 2013: 10).

En esta idea pueden distinguirse dos tesis. La primera es que los motivos no se reflejan en la acción y son meros pensamientos. Así, los motivos se asimilan más a las representaciones oníricas que a otro tipo de estados mentales que generan compromiso práctico. La segunda tesis afirma que un Estado liberal no puede castigar a las personas solo por sus pensamientos porque estos no generan daños a otros. Más allá de la dificultad de articular esta segunda tesis[8], en mi opinión, el problema central del argumento está en la exageración o deformación que implica sostener que los motivos no se reflejan en la acción.

Afirmar que los motivos emocionales son solo pensamientos es una mala comprensión del papel que poseen los motivos en el razonamiento práctico. Hacemos ciertas cosas porque creemos que tenemos razones para hacerlas. Por ejemplo, voy a visitar a mi abuela enferma no solo porque deseo ver a mi abuela sino porque acepto que se debe cuidar a los enfermos. Estos son los motivos que explican mi acción, es decir, que determinan aquello que hice.

Por supuesto, en ocasiones tengo motivos para visitar a mi abuela y no lo hago (por ejemplo, porque me vence la pereza) y la intensidad del

---

8 Creo que esta es una exageración por la manera en que se entienden las obligaciones liberales. Como cuestión de hecho el derecho castiga comportamientos que no producen ningún daño, como la tenencia de armas de guerra, la conducción peligrosa o, incluso, la tentativa. Como paradigma para ver si ciertas normas están justificadas creo que es relevante, pero no puede entenderse que acciones generan o no daño sin incorporar las razones que tenía el agente para hacer lo que hizo. Para un análisis de diferentes maneras de entender el principio de daño en el ámbito de los delitos de odio, véase KAHAN, 2001: 175-193.
También, aunque limitando su argumento a las justificaciones, puede verse GARDNER, 2012: 139-142.

motivo no fue suficiente para determinar mi conducta. Sin embargo, en el ámbito del derecho penal, las acciones relevantes son seleccionadas por los sistemas jurídicos y, por ello, se analizan los motivos siempre que ellos se materialicen en una cierta conducta relevante. Una vez que decidí, por ejemplo, matar a una persona, esta decisión ingresa en el ámbito jurídico, y es aquí donde cobran importancia las razones que tenía para hacer lo que hice. En ese contexto, es importante determinar si el agente ha actuado por ira, por miedo, en legítima defensa, por venganza o por odio. Podemos debatir acerca de por qué el derecho ha otorgado relevancia a la ira (en la defensa de provocación o emoción violenta), al odio (en los homicidios calificados) y no, por ejemplo, a otros motivos emocionales como la piedad. Sin embargo, esta es una discusión normativa acerca de por qué deberían contar ciertos motivos en lugar de otros, y no debe confundirse esa cuestión con una discusión conceptual sobre la manera en que impactan los motivos en la conducta del agente.

Peralta, siguiendo a von Liszt, sostiene que todo motivo posee dos características: la fuerza y la calidad moral. Por tanto, ¿no podría, entonces, también concluirse que cuando se evalúa la culpabilidad solo se puede tener en cuenta la fuerza impulsora del motivo, pero nunca su calidad moral? De otro modo, al tenerlos en cuenta, se abandonaría el principio de responsabilidad por los hechos y se abriría la puerta a castigar estados mentales (Peralta, 2013: 9-10).

Sin embargo, no parece intuitivo reconocer que un motivo posee dos características y, al mismo tiempo, restringir su relevancia a uno de ellos, o limitarse a señalar que uno de ellos es importante únicamente en la esfera moral, pero carece de relevancia para la responsabilidad penal. En parte, el atractivo del argumento criticado surge de la identificación de los estados mentales con cuestiones pasionales, fenomenológicas, que nos suceden, como sensaciones de dolor o placer, y no con una idea más amplia de estados mentales que abarcan objetos intensionales (*i.e.*, dotados de sentido). Si se asume que los motivos se asemejan a las sensaciones, se debe argumentar específicamente al respecto ya que del hecho de catalogarlos como estados mentales no se sigue que no puedan ser evaluados. Más bien, en general, ocurre lo contrario. Es decir, los deseos, las creencias, las emociones son estados mentales y son constantemente objeto de nuestras evaluaciones. Por ejemplo, reaccionamos con preocupación cuando se toma una decisión

basada en creencias arbitrarias, o sentimos indignación ante los prejuicios raciales, etc. Para dar cuenta del papel que juegan los motivos en el razonamiento práctico y en la responsabilidad ellos deben evaluarse con el doble baremo: la fuerza (impulso) y la calidad (razón)[9].

En síntesis, al reprochar por motivos como estados mentales no se evalúa la acción *y después* al pensamiento, sino que no se puede entender qué es aquello que hace el agente si no analizamos los motivos que tenía (o creía tener) para hacer lo que hizo. El segmento de conducta que pretendemos describir puede ser más o menos amplio, y lo que puede contar como motivo en un análisis más estrecho luego puede transformase en una intención o compromiso práctico (GUIBOURG, 1987: 55). Todo depende del grado de especificidad con que pretendamos explicar la conducta. Así, si escucho el timbre y me levanto a abrir la puerta, mi conducta puede explicarse porque deseo abrir la puerta. Pero si me preguntan en un modo más amplio por qué abrí la puerta, podría decir que es porque deseaba dejar pasar a mi amigo, y así sucesivamente (VON WRIGHT, 1988: 83-96). Sin embargo, a pesar de que ciertos motivos cobren relevancia según la descripción de la acción que me interese realizar, ello no quiere decir que solo dependen de cómo un evaluador externo describa la acción (VON WRIGHT, 1983: 53-55). Para poder explicar correctamente una conducta, primero hay que entender qué hizo el agente, y para ello debemos identificar qué quería realizar el agente. Esto solo se hace descubriendo (verdaderamente) qué es aquello que quería realizar el agente[10].

2.2. CRÍTICAS A LAS TESIS PROSPECTIVA Y RETROSPECTIVA

El argumento elaborado anteriormente, por sí solo, alcanza para mostrar que no hay necesidad de elaborar una tesis alternativa a los motivos de

---

9 Elaboraré este argumento más adelante.

10 Un argumento en contra de reprochar los motivos como estados mentales es el que sostiene que existen mayores dificultades probatorias. No trato aquí esta crítica, sin embargo, mi intuición es que las dificultades son las mismas que las de probar cualquier otro estado mental, y si nos importa responsabilizar a los agentes por lo que hacen, el Estado debe realizar el esfuerzo de probar efectivamente aquello que el agente quería hacer.

odio como estados mentales del agente. Esta conclusión sugiere que la tesis positiva de Peralta (*tesis 3*) carecería de utilidad. Sin embargo, esa tesis merece un análisis independiente.

Como se mencionó con anterioridad, la tesis del sometimiento posee dos inconvenientes: es explicativamente ineficiente y es justificatoriamente inadecuada.

En primer lugar, es explicativamente ineficiente, *i.e.*, no sirve para dar cuenta del modo en que dividimos los delitos de odio de otro tipos de delitos. Hay casos paradigmáticos de delitos de odio que no pueden explicarse bajo la noción de sometimiento ya que el autor no pretende que la víctima viva bajo sus reglas. Más bien, el autor pretende que determinadas personas desaparezcan porque las odia. El ejemplo más claro de ello es el del nazi convencido que posee un odio visceral hacia la raza judía. Supongamos que esta persona mata a un judío porque quiere la erradicación de las personas de esa clase. No quiere que el judío viva de un modo diferente, ni siquiera pretende que el judío viva escondiendo su origen y su religión, solo quiere que desaparezca. Lo mismo podría decirse de los responsables del genocidio de Ruanda cometido por el gobierno hutu hacia los tutsi que en cuatro meses exterminó al 75% de la población de esta etnia. Algo similar sucede en casos paradigmáticos de odio a la identidad sexual o de género. En ellos, el autor no desea que el sujeto viva de un modo distinto, desea que no exista. Por ejemplo, con respecto a la identidad sexual o de género, en 2016 la Audiencia Provincial de Barcelona condenó a 12 años de prisión a Abdelhakim E. H. por tentativa de homicidio agravada. El acusado, junto a un menor, agredió a una persona en Sitges. Los insultos comenzaron en un local de ocio y luego, de casualidad, los agresores se encontraron con la víctima en la estación de trenes. Primero la golpearon en la cabeza varias veces, y luego la arrojaron a las vías del tren. Mientras hacían todo ello, además de insultarla, le decían cosas como "*Te vamos a matar con piedras como matamos a los maricones en nuestro país".*

Peralta podría responder que no es importante explicar un caso paradigmático de odio porque él está proponiendo otro concepto de odio. Sin embargo, ello no es así. El objetivo de su trabajo es mostrar por qué la reforma realizada en Argentina que incorpora ciertos delitos de odio está justificada. Al tratar de explicar el reproche más grave de ciertas conductas realizadas con odio de género, a la orientación sexual, etc., Peralta descuida

la idea central de *odio.* Es decir, la noción de sometimiento parece reflejar, en su caso, una preocupación por la violencia machista que oprime a las mujeres, pero la exageración se encuentra en asumir que todos los hechos realizados con odio son de la misma manera.

Peralta parece advertir este problema, y por esa razón en la segunda parte de su fundamentación cobra tanta relevancia la noción de ausencia de provocación. Al final de su artículo, la noción de falta de provocación es la que termina haciendo el trabajo explicativo y justificativo. Explica los agravantes en general y no solo los de odio así como también explica los delitos de odio que no pueden ser explicados con el sometimiento.

> Si la diferencia entre el homicidio simple y el calificado depende de si la víctima es en parte responsable de lo que le sucede, otras constelaciones de casos también quedan abarcadas. Para empezar, podría aplicarse a posibles supuestos de homicidios por odio que, en rigor, no responsen a la idea de sometimiento (Peralta, 2013: 21-22).

Sin embargo, la noción de ausencia de provocación es igualmente deficiente para explicar el modo en que separamos los homicidios más graves (incluidos los de odio) de los homicidios comunes. Transforma casos típicos de homicidios simples como, por ejemplo, casos paradigmáticos de dolo eventual en casos de homicidios agravados. Así, quien juega a la ruleta rusa con un arma cargada, quien hace picadas en la autopista, o incluso el clásico caso de un tirador inexperto que dispara para hacer blanco en una manzana que está sobre la cabeza de quien termina siendo la víctima, se transformarían en homicidio agravado[11]. La consecuencia de ello es una pérdida de equilibrio conceptual que transforma por completo las nociones de dolo, intención y compromiso práctico. Incluso pierde el trasfondo justificatorio que Peralta también aceptaría, *i.e.*, que el paradigma del reproche son las acciones intencionales.

---

11 En otro lugar he criticado que estos casos sean identificados como dolosos, por lo que con mayor razón insistiría en no catalogarlos como homicidio calificado. Véase, por ejemplo, Manrique, 2012: 107 ss.

Recordemos que el esfuerzo de Peralta por justificar el homicidio por odio se centra en dos ideas intrínsecamente relacionadas. Por una parte, hay una intuición compartida (y correcta) de que los homicidios por odio son más graves que los homicidios 'comunes'. Por otra parte, no puede entenderse al odio como un estado mental porque ello genera que reprochemos los pensamientos y esto no está justificado en un derecho penal liberal. Creo que incluso si asumimos como adecuada la fundamentación que Peralta propone, entonces todavía podrían indicarse como problemáticas las consecuencias más bien extrañas que su posición genera. Los potenciales autores podríamos pensar que, si el coste de tener un derecho penal de hecho es que todos los casos de homicidio que podamos cometer se transformarían en homicidio agravado, entonces es mejor quedarse con un derecho que castigue 'solo ciertos pensamientos', incluso aunque desde un punto de vista liberal (equivocadamente) se afirme que es derecho penal de autor.

Además, si existe provocación no se aplicaría la agravante a casos que uno podría decir más graves. Así, el hecho de que me hubiese peleado, o hubiese incluso insultado a la madre de mi asesino, no justifica que el homicidio sea menos grave si, finalmente, el autor decide matarme, porque soy judía, o porque soy homosexual. Pero, incluso los casos de homicidios que resultan de la violencia doméstica pueden catalogarse, en algunas ocasiones, como casos de 'escalada'. Esto es, casos donde se comienza por un conflicto, que puede ser trivial, y se termina en la muerte de uno de los cónyuges, casi siempre la mujer. Estos casos quedarían fuera de la justificación de los delitos agravados, y sin embargo son, en gran medida, casos a los que Peralta pretende abarcar (esto dejando de lado que en sistemas jurídicos como el argentino se agrava el homicidio en caso de existir violencia de género).

En segundo lugar, creo que la noción de ausencia de provocación es inadecuada para justificar el mayor reproche. No hay manera de entender un conflicto sin buscar la razón del mismo. Esto nos hace preguntar *por qué*, y ello significa preguntarse por las razones que tenía el agente para hacer lo que hizo. Esto no es más que otro modo de preguntar sobre sus intenciones y también sobre los motivos. Sin embargo, en lugar de averiguar los motivos del homicidio, lo central se traslada al comportamiento de la víctima y los momentos anteriores al homicidio. La pregunta es, si

estamos evaluando el homicidio, ¿por qué un derecho penal liberal está justificado en preguntar sobre las razones o motivos que tenía la víctima para actuar como lo hizo en un momento anterior al homicidio y no puede cuestionarse sobre las razones que tuvo el agente para hacer lo que hizo en el momento en que cometió el delito?

En conclusión, si Peralta tiene razón sobre los problemas que presenta entender los motivos como estados mentales, estos mismos problemas se trasladan a su posición, ya que, si pretende entender lo que pasó, en tal caso debe analizar las razones del conflicto. "Los homicidios no ocurren *sin motivos*, sino que suelen ser la consecuencia de un conflicto" (Peralta, 2013: 17-18). Pero, si Peralta no tiene razón en la crítica a los motivos como estados mentales, entonces la tesis del sometimiento y la ausencia de provocación no solo son deficientes sino también innecesarias.

Conclusión. Un balance sobre los motivos de odio y el sometimiento

En este trabajo he distinguido dos problemas centrales que se identifican en el debate del odio como motivo para agravar la responsabilidad. En primer lugar, con respecto a los problemas de legitimidad, he identificado una manera de responder a quienes afirman que reprochar por los motivos es castigar los pensamientos. En segundo lugar, en cuanto a los problemas de fundamentación, he analizado una manera novedosa de justificar que el reproche por los motivos está justificado, pero su fundamento no se encuentra en el valor que poseen los estados mentales, sino que se centra en la noción de sometimiento.

Pero, como he intentado mostrar anteriormente, la concepción del sometimiento no está exenta de problemas. Primero, he tratado de mostrar dos tipos de inconveniente. Sostuve que como cuestión explicativa es ineficiente, *i.e.*, no muestra el modo en que distinguimos o en que los códigos distinguen entre delitos de odio y otro tipo de delitos. Segundo, a nivel justificatorio no es adecuada, y no lo es porque, o bien vuelve a recurrir al odio como estado mental, o bien el precio a pagar por la concepción del sometimiento es demasiado elevado y transforma en agravados una gran variedad de delitos que no queremos agravar. La consecuencia de ello es una pérdida de equilibrio conceptual de las nociones de homicidio agravado.

En conclusión, creo que mientras más fina sea la explicación de por qué el agente actuó de determinada manera, en mejores condiciones nos encontramos de imponer, en su caso, un castigo más adecuado a la conducta efectivamente realizada. Por supuesto, el derecho no le otorga relevancia a todos los motivos, sino solo a los que considera importantes (para mal o para bien), y al tenerlos en cuenta nos hace identificar mejor cómo castigar[12].

Los motivos como el odio hacia alguna raza o algún género han sido tomados como agravantes de ciertas conductas; en este sentido, hacen pensar que ellos solo reflejan a un autor más culpable que realizó la misma conducta que otro, *e.g.*, homicidio, que no poseía ese motivo, pero ello no tiene que ser necesariamente el caso. En ocasiones los motivos definen aquello que el agente hace, *e.g.*, tortura, defensa propia, estado de necesidad, etc., y no hay necesidad en entender que ellos cumplen una sola función en la descripción de la acción[13].

REFERENCIAS

Díaz López, J. A., 2013: *El odio discriminatorio como agravante penal. Sentido y alcance del artículo 22.4 del Código Penal.* Madrid: Civitas.

Gardner, J., 2007: *Offences and Defences.* Oxford: Oxford University Press. Citado por la traducción de Manrique-Peralta, *Ofensas y defensas. Ensayos selectos sobre filosofía del derecho penal.* Madrid-Barcelona-Buenos Aires: Marcial Pons, 2012.

Guibourg, R., 1987: *El fenómeno normativo.* Buenos Aires: Astrea.

---

12 La discusión sobre si el Estado eligió bien los motivos que considera relevantes es una discusión política que no daré aquí. Sin embargo, creo que esta es la discusión más importante que debe darse en el ámbito del reproche de los motivos de odio. La crítica conceptual, que he tratado de mostrar que es equivocada, solo ha servido para ocultar esta discusión.

13 En el mismo sentido, véase Mathis, 2018: 10. En términos dogmáticos estoy afirmando que los motivos no muestran solo cuán culpable es el agente, sino que en ocasiones los motivos impactan en el injusto (el hecho). Ellos definen qué hizo el agente. El hecho de que la teoría del delito no encuentra una categoría clara donde colocar los motivos no es un problema para evaluar los motivos, sino que es algo vinculado a las categorías que hemos establecido para analizar el delito.

González Lagier, D., 2009: *Emociones, responsabilidad y derecho.* Barcelona: Marcial Pons.

Hurd, H. y Moore, M., 2004: "Punishing hatred and prejudice", *Standford Law Review*, 56 (5): 1081-1146.

Kahan, D. y Nussbaum, M., 1996: "Two conceptions of emotion in criminal law", *Columbia Law Review*, 96 (2): 270-372.

Kahan, D., 2001: "Two liberal fallacies in the hate crime debate", *Law and Philosophy*, 20: 175-193.

Manrique, M. L., 2012: *Acción, dolo eventual y doble efecto.* Barcelona: Marcial Pons.

Mañalich, J. P., 2016: "Arrebato y obcecación pasionalmente condicionados como atenuantes por un homicidio frustrado", *Revista de Estudios de Justicia,* Universidad de Chile, 25.

Mathis, S., 2018: "Motive, action, and confusions in the debate over hate crime legislation", *Criminal Justice Ethics,* 37 (1): 1-20.

Miró Linares, F. (dir.), 2017: *Cometer delitos en 140 caracteres. El derecho penal ante el odio y la radicalización en Internet.* Madrid-Barcelona-Buenos Aires-Sao Paulo: Marcial Pons.

Nino, C. S., 1980: *Los límites de la responsabilidad penal.* Buenos Aires: Astrea.

Peralta, J. M., 2011: "El 'feminicidio' y el derecho penal", *Comercio y Justicia*, disponible en: https://comercioyjusticia.info/blog/opinion/el-feminicidio-y-el-derecho-penal/

Peralta, J. M., 2012: *Motivos reprochables. Una investigación acerca de la relevancia de las motivaciones individuales para el derecho penal liberal.* Madrid: Marcial Pons.

Peralta, J. M., 2013: "Homicidio por odio como delito de sometimiento", *InDret,* 4. Disponible en: http://www.indret.com/pdf/1005.pdf

PÉREZ MANZANO, M., 2018: "La caracterización del femicidio de la pareja o expareja y los delitos de odio discriminatorio", *Revista de la Facultad de Derecho de PUPC*, 81: 163-196.

TADROS, V., 2013: "Wrongdoing and Motivation", en DUFF, R. A. y GREEN, S. (eds.), *Philosophical Foundations of Criminal Law*. Oxford, Oxford University Press, 206-227.

VON WRIGHT, G. H., 1981: "Explanation and understanding of action", *Revue Internationale de Philosophie*, 35. Citado por la reimp. en *Practical Reason Philosophical Papers I*, Ithaca-New York: Cornell University Press, 1983, 53-66.

VON WRIGHT, G. H., 1988: "An essay on door-knocking", *Rechtstheorie*, 19: 275-288. Citado por la reimp. en *In the Shadow of Descartes. Essays in the Philosophy of Mind*. London, Kluwer Academic Publishers, 1998, 83-96.

WILENMANN, J., 2016: "El sistema de graduación de la pena del homicidio en el derecho chileno", *Política Criminal*, 11 (22): 721-765. Disponible en: http://www.politicacriminal.cl/Vol_11/n_22/Vol11N22A12.pdf

Editado por el Departamento de Publicaciones
de la Universidad Externado de Colombia
en julio de 2021

Se compuso en caracteres Ehrhardt Regular de 12 puntos
y se imprimió sobre papel Holmen Book Cream de 60 gramos
Bogotá (Colombia)

Post tenebras spero lucem